Rapid visual form-based processing of (some) grammatical features in parallel reading: An EEG study in English

Donald Dunagan[1]*, Tyson Jordan[1], John T. Hale[1], Liina Pylkkänen[2,3], & Dustin A. Chacón[1]

[1]Department of Linguistics, University of Georgia
[2]Department of Psychology, New York University
[3]Department of Linguistics, New York University

*Corresponding author

Donald Dunagan
142 Gilbert Hall
University of Georgia
Athens, GA 30602-6205
dgd45125@uga.edu

Abstract

Theories of language processing – and typical experimental methodologies – emphasize word-by-word processing of sentences. This paradigm is good for approximating speech or careful text reading, but arguably, not for the common, cursory glances used while reading short sentences (e.g., cellphone notifications, social media posts). How much grammatical information can be gleaned from a single glance? In an electroencephalography (EEG) study, brain responses to grammatical (*the dogs chase a ball*) stimuli diverged from scrambled counterparts (*a dogs chase ball the*) ~300ms post-sentence onset, and from non-lexical consonant strings (*thj rjxb zkhtb w lhct*) ~220ms post-sentence onset. This demonstrates early recognition and cursory analysis of linguistic stimuli. However, EEG responses do not diverge between grammatical sentences and their counterparts with ungrammatical agreement (*the dogs chases a ball*). Additionally, the surprisal of individual words affects the EEG signal at non-uniform time points, from 250ms–600ms. We propose that, in cursory reading in a single glance, readers extract some sentence-level information, such as basic syntactic structure, then 'fill in' some lexical details in a top-down fashion afterwards. This cursory syntactic analysis, however, is not detailed enough to support detection of formal syntactic agreement errors. We suggest this may be due to either the minimal visual salience of agreement morphology in English (-*s*), or a strategic ignoring of semantically-inert syntactic features for the sake of extracting a semantic 'gist.'

Keywords. sentence processing, syntax, agreement, reading, EEG, psycholinguistics

1. Introduction

Most theories of language processing assume incremental word-by-word processing (Frazier & Fodor 1978; Frazier 1987; Hale 2001; Lewis & Vasishth 2005; Brasoveanu & Dotlačil 2020). In the cognitive neuroscience of language, standard methodology follows this general framework – sentences are presented serially, word-by-word, with evoked brain responses measured at each word. This is motivated by the fact that speech, sign, and careful reading of longer texts are serial. However, seriality is a poor model for an increasingly common mode of reading: casual glances at short sentences (e.g., TikTok videos captions, text notifications). Our question is: How much linguistic information can be gleaned from a single fixation, and how quickly?

Standard electrophysiological models of word-by-word sentence processing (e.g., Friederici 2011; Friederici & Kotz 2003; Kaan 2007) ascribe a number of different event-related potentials (ERPs) to different hypothesized aspects of language processing. The early left anterior negativity (eLAN; generally observed over left anterior electrodes at 100–200ms) is associated with automatic local structure building, and can be found for word category violations (see Friederici & Weissenborn 2007). A later left anterior negativity (LAN; at 300–500ms) is associated with case marking and morphosyntactic agreement (e.g., subject-verb number disagreement; see Molinaro et al. 2011). The N400 (a negative-going centro-parietal component around 400ms) is associated with the access of lexical-semantic information (Lau et al. 2008; Friederici 2011) and is highly correlated with surprisal (e.g., Frank et al. 2015; Lowder et al. 2018). Lastly, The P600 (a positive-going centro-parietal component between 500–900ms) is associated with the integration and retrieval of sentence-level syntactic and semantic information (Friederici 2011). These ERPs – and their interpretations – importantly, are contextualized by word-by-word language comprehension; that is, the processing of lexical items given a partial syntactic analysis. It is unclear to what degree these commonly-observed responses from serial reading and listening paradigms should be evoked in the reading of parallel input. This is because serial processing involves resolving a new word with a prior context word-by-word, whereas the psycholinguistic mechanisms used in parallel reading may not be in phase with words.

Several threads of research demonstrate that the visual system is capable of extracting information from more than a single word in one fixation. For example, skilled English readers are capable of detecting several letters to the right of a fixation (Rayner 1998) and access some features of words in the parafovea (Rayner et al. 2003; Schotter et al. 2012; Antúnez et al. 2022). Therefore, it is possible that readers use parallel processing strategies, integrating information from across the visual field (Snell & Grainger 2019). Similarly, findings from single-word processing studies also suggest rapid and parallel processing. A number of results suggest rapid, form-based processing of morphologically complex forms into their constituent parts (Rastle et al. 2000; Rastle et al. 2004; see Taft & Forster 1975). Magnetoencephalography (MEG) recordings show that morphologically complex forms (*re-made*; *sing-er*) exhibit increased activity in occipito-temporal regions around 170ms (the 'M170' event-related field; Zweig & Pylkkänen 2009), and that this activity correlates with the relative probability of the whole word vs. the stem (Lewis et al. 2010; Solomyak & Marantz 2011; Gwilliams & Marantz 2018; see also Holcomb & Grainger 2006; Grainger & Holcomb 2009; Morris et al. 2013). This M170 response temporally precedes neural responses reflecting lexico-semantic interpretation of the word (Pylkkänen & Marantz 2003; Fruchter & Marantz 2015). This suggests that early stages of reading assess features of the form of multiple morphemes in a single fixation in single-word reading, which could in principle extend to short phrasal stimuli as well.

Nonetheless, it remains controversial how much linguistic information can be extracted from a single fixation, and how parallel these processes might be (White et al. 2019; Snell & Grainger 2019). Recent work has explored the visual system's capacity to extract grammatical and semantic information from short sentences read in a single fixation, presented rapidly, only visible for ~200ms. Asano & Yokosawa (2011) found that Japanese readers were more accurate at recalling words in rapidly-presented sentences in which target words were contextually appropriate vs. sentences that had a semantic anomaly. To account for this, they propose that features of the sentence can be processed in parallel, and that a 'gist' can be rapidly extracted. In a similar vein, in a series of studies in French sentences with 200ms presentation times, Snell, Grainger, and colleagues found an increase in word recall accuracy for grammatical and semantically plausible sentences vs. scrambled or implausible sentences (Snell & Grainger 2017; Snell & Grainger 2019; Wen et al. 2019; Massol et al. 2021), which they refer to as the 'sentence superiority effect.' In an electroencephalography (EEG) study, Wen et al. (2019) found that, compared to their scrambled counterparts, grammatical sentences presented in this fashion resulted in a reduction in amplitude of the N400, a component which is inversely correlated with the accessibility of lexical semantics in a context (Kutas & Hilyard 1984; Lau et al. 2008; Bornkessel-Schlesewksy & Schlesewsky 2019). Similar to Asano & Yokosawa (2011), Wen et al. (2019) proposed that some grammatical information can be extracted in parallel from the fleeting stimulus, which can then interactively facilitate lexical access, resulting in a reduction in the N400 component.

Further work using MEG elaborates on this paradigm, reliably finding early responses to grammatical manipulations, often 300ms post-stimulus onset or faster. In an MEG study in English, Fallon & Pylkkänen (2023) found that grammatical sentences elicited greater activation than length-matched noun lists in the left posterior temporal lobe, ~180ms post-sentence onset. Similarly, Flower & Pylkkänen (2024) found sensitivity to grammatical sentences vs. reversed sentences starting at ~210ms post-sentence onset, localized to a broad left fronto-temporal and temporo-parietal language network. However, these two studies also demonstrated that the syntactic information gleaned from a single fixation may not be fully detailed. Fallon & Pylkkänen (2023) showed a similar pattern of activation for English sentences with agreement errors as English sentences without agreement errors. In the same vein, Flower & Pylkkänen (2024) found that brain activity did not distinguish between grammatical sentences and sentences with two-word transpositions until approximately 320ms. Finally, in an MEG study in Danish, Krogh & Pylkkänen (2024) replicated the sentence superiority effect at ~230ms, with activity localized in left inferior frontal regions and left anterior temporal regions. Moreover, they observed that different features of well-formed grammatical sentences elicited neural effects at different times, with argument structure features impacting activity starting at ~250ms, and yes/no questions diverging from declarative sentences at ~500–720ms.

Taken together, these findings suggest a rapid parallel activation of syntactic, semantic, and lexical features from parallel input, often at time scales much earlier than typically reported in studies using more traditional, word-by-word presentation. Although the models proposed by Asano & Yokosawa (2011) and Wen et al. (2019) involve a close interactive relation between sentence-level features and lexical access, the only investigation into lexical processing itself are the analyses on the behavioral responses to the memory recall probe. These occur after the fleeting sentence has been encountered, i.e., after the participant may have reconstructed or 'guessed' the details of the sentence using top-down inferences (cf. Staub 2023). Although Wen et al.'s (2019) N400 results may suggest that lexico-semantic processing facilitates syntactic processing with parallel presentation, it is generally unclear to what extent the interpretation of evoked responses in serial presentation are meant to align with evoked responses in parallel presentation. Thus, it remains to be seen whether properties of individual lexical items exert an effect on the EEG signal at the same time-scale as grammatical features.

Here, we present a novel EEG experiment, building on Wen et al.'s (2019) and Fallon & Pylkkänen's (2023) designs. Our experiment has two questions: Given the sentence superiority effect, does the brain 'notice' sentences with ungrammatical subject-verb agreement, presented in a rapid parallel visual presentation style? The second question is whether the sentence superiority effect coincides with lexical access, as indexed by the correlation between N400 responses and surprisal values of individual lexical items.

In addition to comparing grammatical sentences (Grammatical; *the dogs chase a ball*) vs. scrambled counterparts (Scrambled; *a dogs chase ball the*), we include a non-lexical consonant string (Consonant String; *thj rjxb zkhtb w lhct*) condition and an agreement violation (Ungrammatical; *the dogs chases a ball*) condition. The comparison of the brain responses to the Grammatical and Scrambled conditions permits us to replicate the findings of Wen et al. (2019) and Fallon & Pylkkänen (2023), establishing the time of the 'sentence superiority effect.' The key comparison is between the Grammatical condition and the Ungrammatical condition, which only differ in the subject-verb agreement relation. In standard serial presentation styles, agreement violations exhibit reliable evoked responses, often eliciting an eLAN and a P600 (Molinaro et al. 2011, among many others). Comparison of the brain responses to the Consonant String and Grammatical conditions allows for isolating brain activity that distinguishes between familiar and interpretable stimuli with visually unfamiliar and uninterpretable stimuli. Finally, we also conducted single-trial analyses to investigate the effect of individual lexical items on the evoked response. For sentences with serial presentation, surprisal – how 'surprising' it is to observe a word in a context – is highly correlated with N400 amplitudes (e.g., Frank et al. 2015; Lowder et al. 2018).

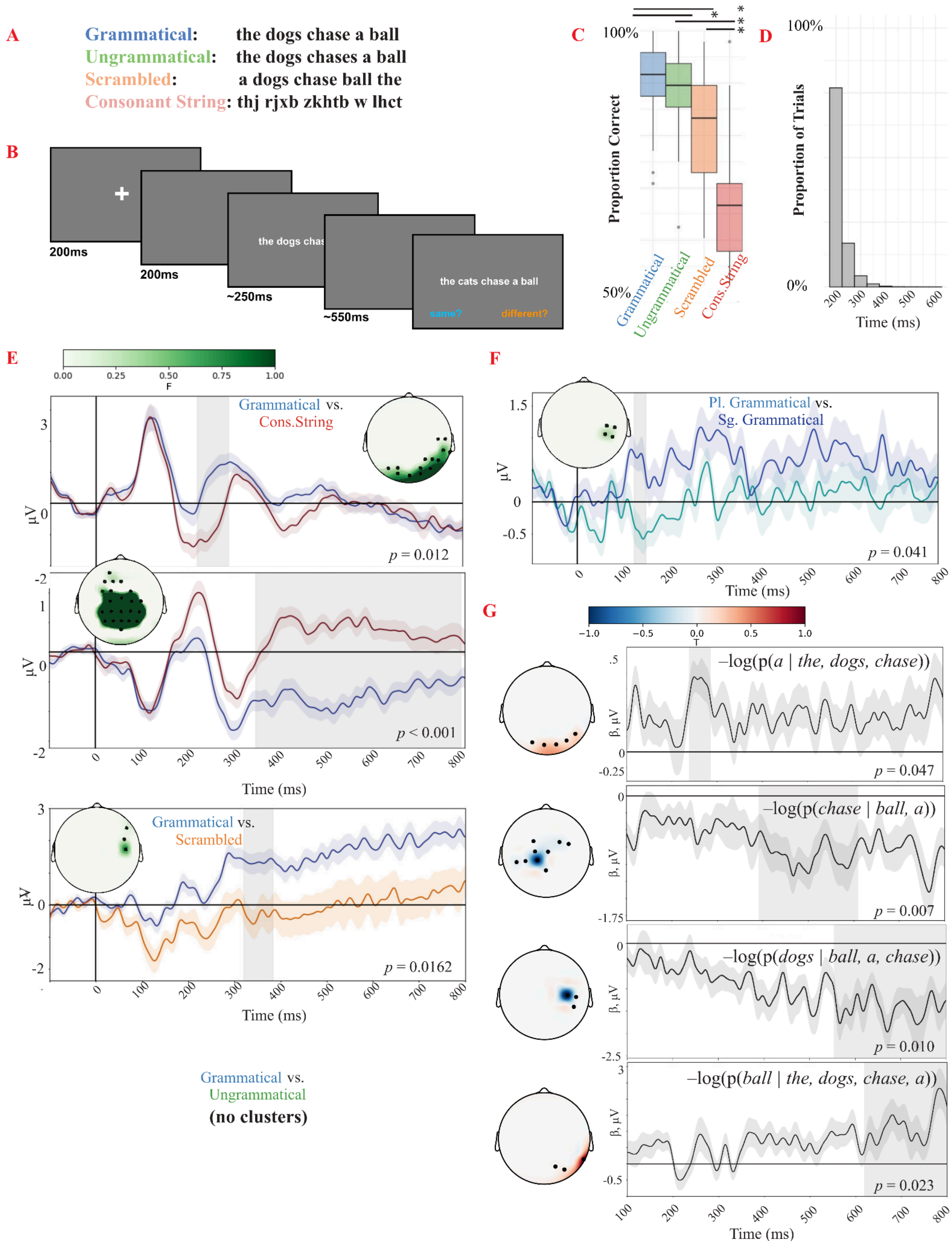
A
Grammatical: the dogs chase a ball
Ungrammatical: the dogs chases a ball
Scrambled: a dogs chase ball the
Consonant String: thj rjxb zkhtb w lhct
B
200ms
200ms
the dogs chas
~250ms
~550ms
the cats chase a ball
same?
different?
C
100%
50%
Proportion Correct
Grammatical
Ungrammatical
Scrambled
Cons.String
D
100%
0%
Proportion of Trials
Time (ms)
E
0.00 0.25 0.50 0.75 1.00
F
Grammatical vs.
Cons.String
p = 0.012
p < 0.001
Time (ms)
Grammatical vs.
Scrambled
p = 0.0162
Time (ms)
Grammatical vs.
Ungrammatical
(no clusters)
F
Pl. Grammatical vs.
Sg. Grammatical
p = 0.041
Time (ms)
G
−1.0 −0.5 0.0 0.5 1.0
T
–log(p(a | the, dogs, chase))
p = 0.047
–log(p(chase | ball, a))
p = 0.007
–log(p(dogs | ball, a, chase))
p = 0.010
–log(p(ball | the, dogs, chase, a))
p = 0.023
β, μV
Time (ms)

Figure 1. Experiment structure and behavioral and EEG results. (A) Example stimulus items for each condition. (B) Trial structure. (C) Accuracy by condition on the match/mismatch task. Starred pairwise comparisons are computed from pairwise comparisons of the logistic regression fit in Table 1. Each comparison is significant ($p < 0.001$), except for the comparison between Grammatical and Ungrammatical. (D) Proportion of trials by presentation time. 77% of trials were at the ceiling presentation time of 200ms. (E) Significant clusters from one-way ANOVA analyses. Topographic plots show the distribution of F-values and significant sensors included in each cluster. Time series show the average activation in these sensors. Gray shading corresponds to the temporal extent of the cluster. We report on the largest and earliest cluster in the Grammatical vs. Consonant String analysis, and the only significant cluster for the Grammatical vs. Scrambled analysis. (F) The single significant cluster from the *post hoc* one-way ANOVA analysis comparing singular and plural Grammatical trials. (G) Significant clusters from the single-trial analysis. Topographic plots show the distribution of T-values and significant sensors included in each cluster. Time series show the average β coefficient value in these sensors. Gray shading corresponds to the temporal extent of the cluster.

2. Experiment

There were three primary goals of this experiment. The first was to replicate the sentence superiority effect observed in both EEG (Wen et al. 2019) and MEG (Fallon & Pylkkänen 2023; Flower & Pylkkänen 2024; Krogh & Pylkkänen 2024), i.e., a divergence in brain responses for grammatical sentences vs. a non-sentence control. Like previous studies, we compare simple transitive sentences (*the dogs chase a ball)* against non-sentence controls. In our study, we use both scrambled sentences (*a dogs chase ball the*), in which words occur in an order inconsistent with English grammar rules, and unpronounceable consonant strings that are matched in length to the grammatical controls (*thj rjxb zkhtb w lhct*). The scrambled sentences are included to allow for a comparison to previously published results. The addition of the consonant string controls allows us to determine to what extent the sentence superiority effect response corresponds to detecting linguistic information vs. detecting familiar visual stimuli. If the difference in the brain response to grammatical sentences vs. scrambled sentences is similar to the brain response to grammatical sentences vs. consonant strings, then the sentence superiority effect may reduce to properties of recognizing some well-rehearsed visual stimuli, such as high frequency *n*-grams and short 'sight words.' By contrast, if the sentence superiority effect is suggestive of grammatical analysis, as assumed by previous studies, then we predict that the grammatical vs. scrambled comparison should yield a different pattern of responses than the grammatical vs. consonant string comparison.

The second goal was to determine whether the neural responses to sentences displayed in the rapid parallel presentation style would be sensitive to subject-verb agreement errors. Subject-verb agreement errors are observed to elicit very strong biphasic responses in EEG recordings (a LAN followed by a P600) in serial reading experiments, which suggests both detection of the ungrammaticality followed by a repair attempt (Angrilli et al. 2002; Coulson et al. 1998, Osterhout & Mobley 1995; see Molinaro et al. 2011, for a review). However, Fallon & Pylkkänen (2023) and Flower & Pylkkänen's (2024) MEG results suggest that at least in the early stages of processing, a neural sentence superiority effect is obtained even for sentences containing agreement violations and word order transpositions. We sought to determine whether this failure to detect ungrammatical agreement forms would also be observed in EEG.

Finally, the third goal was to conduct an exploratory, single-trial analysis to determine when and whether individual lexical items affected the neural response. Crucially, the model presented by Wen et al. (2019) suggests that lexical activation is facilitated by early processing of the syntactic structure, assuming that their N400 response reflected such facilitated lexical activation. However, all previous results have focused on comparing the neural responses to different syntactic structures. We hypothesized that, if grammatical structure facilitates lexical access in a parallel, cascaded fashion, then the sentence superiority effect should occur at roughly the same time scale as the effect of the surprisal values of the lexical items in the sentence. By contrast, if grammatical features are accessed first, then we may predict later or variable onsets of surprisal effects of different lexical items.

Following Fallon & Pylkkänen (2023), Flower & Pylkkänen (2024), and Krogh & Pylkkänen (2024), we replace the single word memory probe task used by Asano & Yokosawa (2011), Snell & Grainger (2017), and Wen et al. (2019) with a simple match/mismatch task. This task is more suitable for electrophysiological recordings since it involves fewer motor movements, is simpler to implement and easier for participants, and partially circumvents critiques about top-down 'guessing' strategies that may facilitate performance in the original task (Staub 2023).

2 Methods

2.1 Participants

Thirty-five self-identified English speakers were recruited from the ANONYMOUS CITY community. Thirty-three participants were right handed, as assessed by the Edinburgh Handedness inventory (Oldfield 1971). All participants had normal or corrected-to-normal vision, and reported no history of language impairment. Participants gave written informed consent, and were compensated $10 per 30 minutes of participation time. The study was approved by ANONYMOUS UNIVERSITY (#ANONYMOUS NUMBER).

2.2 Materials

We prepared 50 sets of sentences in 4 separate conditions – Grammatical, Scrambled, Ungrammatical, and Consonant String (Fig. 1A). All Grammatical trials were simple transitive sentences, consisting of five words. These sentences were a determiner, an animate noun subject, a transitive verb, a different determiner, and an inanimate noun object. Half of the grammatical trials consisted of a plural subject noun ending in -*s* and a plural verb; the other half had a singular subject noun and a singular verb ending in -*s*. All words were monosyllabic and monomorphemic, with the exception of the plural morphemes -*s*.

We also generated length-matched unpronounceable consonant strings. These were included as a non-lexical control, to delineate between early brain responses that might index early linguistic analysis (i.e., Grammatical vs. Ungrammatical, Scrambled) and early responses that might index initial detection of familiar visual forms that may not necessarily index any linguistic analysis (Grammatical vs. Consonant String).

Scrambled sentences were constructed by randomly permuting the words of the Grammatical trials, and then manual verification that the new string also did not form a new grammatical sentence. We did not exclude scrambled phrases that could create local coherent sub-strings (e.g., the relative clause interpretation of *a ball the dogs chase*).

Ungrammatical trials were created by either removing the required singular verb agreement morpheme -*s* on the verb (*the dog chase a ball*), or adding an unlicensed singular verb agreement morpheme -*s* on the verb (*the dogs chases a ball*) from the Grammatical trial. Because number was counterbalanced across item sets, length was matched between the Grammatical and Ungrammatical conditions.

Mismatch trials were constructed by replacing either the subject noun, the verb, or the object noun with a semantically similar length-matched foil (*the dogs chase a ball* ~ *the cats chase a ball*). The probes in the match/mismatch task maintained all other features, i.e., scrambled word order, ungrammatical agreement, or consonant strings. Half of the stimuli were paired with a match trial, and the other half were paired with a mismatch trial. This pairing was done within-stimulus set, so participants could not leverage the identity of specific lexical items to guess whether a trial would be a match/mismatch. Instead, participants must attend to the two nouns and verbs, i.e., the three content words.

2.3 Procedure

We used an adapted version of a match/mismatch task employed in prior studies on behavioral and neural sentence superiority effects (Pegado & Grainger 2020; Pegado et al. 2021; Fallon & Pylkkänen 2023; Flower & Pylkkänen 2024, Krogh & Pylkkänen 2024). In Snell & Grainger's studies (2017, 2019), sentences were displayed for 200ms, followed by a memory probe task. Participants were asked to recall a word in a cued position on the screen using a visual indicator, and they provided their responses by typing. This task is less suitable for electrophysiological recordings, in which motor movements can elicit strong electromagnetic noise. Instead, we replace the cued word recall task with a simple match/mismatch task, in which participants determine whether an untimed, second sentence is the same as or different from the rapidly displayed sentence.

All sentences were displayed centered on a screen, printed in a white monospace font against a dark gray background. Target sentences were displayed for a variable length, followed by a dark gray blank screen of variable length, which was adjusted such that the total time of the target sentence and the blank screen summed to 800ms. After the blank screen, participants saw another short sentence, which remained visible until participants pressed a button to indicate whether it matched. Participants entered their response using a keyboard, with the 'f' key corresponding to 'match' and the 'j' key corresponding to 'mismatch.' On-screen reminders ('same?', 'different?') were presented at each memory probe, to ensure participants remembered the mapping between key presses and responses. Participants were kept at a uniform distance from the screen, approximately 70cm from nasion to center of the screen. The visual angle subtended of the stimuli was approximately 15 degrees. The example trial structure is shown in Fig. 1B.

Half of the trials were 'Mismatch' trials, and the other half were 'Match.' Unlike in prior RPVP studies using a match task (Pegado & Grainger 2020; Pegado et al. 2021; Fallon & Pylkkänen 2023; Flower & Pylkkänen 2024, Krogh & Pylkkänen 2024), sentence and blank screen display time were dynamically adjusted based on participant's performance. We implemented this variable display time mechanic in part to conduct parallel experiments in other languages with different writing systems. We did not know *a priori* whether the minimum feasible stimulus length would be the same across writing systems and languages, and we expected by-participant variance as well. Thus, we wanted the trial length to be adjusted dynamically to identify the ideal trial length for each subject. Sentence display time was constrained to vary between 200 and 600ms, initialized at 200ms. Sentence display time increased by 50ms after incorrect responses, and decreased by 50ms after correct responses. Participants received feedback on incorrect trials, with a 1000ms delay before the next trial. Faster presentation times were necessary to ensure that participants could not saccade during the critical time window, and thus read the stimuli serially. For our experiment, responses were remarkably accurate across trials (see Results and Fig. 1C). Thus, the average stimulus length was 216ms (SE = 0.43ms), and 77.0% of trials were at the minimum speed of 200ms, suggesting that variable display time should not be a substantial influence on performance or EEG responses (Fig. 1D).

Stimuli were presented in 4 separate blocks. Participants saw one trial per item set per block, with conditions and items evenly distributed between the 4 blocks. After each block, participants were instructed to take a break, which ranged from 2-5 minutes, and were provided with jokes or photos of animals to encourage them to not progress immediately into the next block. An experimenter was present with the participant during the recording, and was available to answer questions or assist the participant if needed.

EEG signals were recorded using a 64 channel Ag/Cl BrainVision actiChamp+ system (Gilching, Germany). Impedance of the EEG sensors was reduced by the application of SuperVisc gel, and lowered to <25kΩ. On-line EEG recording was referenced to FCz according to manufacturer standards, and then re-referenced to average sensors offline. Participants engaged in an unrelated task that is not reported here. The order of the two tasks was counterbalanced. Participants also engaged in a series of localizer tasks at the beginning of each recording session, but we do not report this here.

3. Results

3.1 Behavioral Results

All participants responded with >50% accuracy on all conditions. No participants were excluded on the basis of behavioral data. The mean results are graphed in Fig. 1C.

For analysis, we fit a logistic regression model, with correct response as the dependent variable, with condition as a 4-level factor, and with participant and item as random effects. More complex random effect structures failed to converge. The condition factor was fit with Grammatical as the intercept, since the research question is how each condition diverged from the Grammatical condition. The results of the model are presented in Table 1. Afterwards, we conducted pairwise comparisons to assess which conditions diverged. Pairwise comparisons were conducted on the logistic regression model, with Tukey HSD correction, using the lsmeans package in R. All conditions diverged from each other (z-ratios > 6, ps < 0.001), except for the pairwise comparison between Grammatical and Ungrammatical ($\beta = 0.156$, SE = 0.12, z-ratio = 1.35, $p = 0.53$), strongly suggesting that there is no detectable difference between sentences with grammatical agreement vs. sentences with agreement errors.

	β	SE	*z*	*p*
(Intercept)	2.47	0.15	16.25	**<0.001**
Condition: Ungrammatical	–0.16	0.12	–1.35	0.18
Condition: Scrambled	–0.82	0.11	–7.67	**<0.001**
Condition: Cons. String	–1.72	0.10	–17.01	**<0.001**

Table 1. Results of the logistic regression fit to the behavioral responses. The model formula was Correct ~ Condition + (1|Subject) + (1|Item). Bolded p-values are significant at alpha < 0.05.

3.2 EEG Processing

All EEG preprocessing was conducted in MNE-Python (Gramfort et al. 2013), and statistical analyses were conducted using MNE-Python and Eelbrain (Brodbeck et al. 2023). Raw

EEG data was filtered off-line between 0.1-40Hz, using an IIR bandpass filter. We then removed flat or noisy sensors, and interpolated them. We then re-referenced the EEG data to an average reference, instead of FCz. Afterwards, we extracted epochs from –100ms to 800ms post-stimulus onset, i.e., the time of the rapid sentence presentation and the subsequent blank screen. The 100ms pre-onset period was then used for baseline correction. We then used independent component analysis (ICA) to identify semi-regular endogenous electromagnetic noise sources, including eyeblinks, eye movements, and heartbeats. These components were removed. Following ICA, we then automatically rejected all epochs that exceeded a 100µV peak-to-peak threshold. Afterwards, we visually inspected and removed other problematic epochs.

We then created condition averages for each participant for factorial analyses. Epochs were normalized in number using the default function in MNE-Python before conducting analyses. After exclusion of the 18.4% incorrect trials, unusable epochs, and trial condition normalization, 38.8% of trials were excluded.

3.3 Average Results

There were 4 conditions in our experiment, but the important question was when and whether the Grammatical trials diverged from the three ungrammatical foils: Consonant String, Scrambled, and Ungrammatical. To this end, we conducted three separate one-way ANOVAs. Because we partitioned the same data into 3 separate analyses, we use a Bonferroni-corrected alpha of 0.05/3=0.0167. The ANOVAs were fit for Grammatical vs. Consonant String, Grammatical vs. Scrambled, and Grammatical vs. Ungrammatical. Each ANOVA had Grammatical as the intercept, and a coefficient was fit for the ungrammatical condition. The main results of both average and single-trial analyses (see section 3.4) are summarized in Fig. 1E–G.

We conducted our analyses using spatio-temporal cluster-based permutation tests (Maris & Oostenveld 2007). This is a non-parametric method for first identifying clusters of significant responses that are contiguous in space and time, and then bootstrapping a null distribution from the data to estimate the statistics of the cluster. This allows for correcting for multiple comparisons, while acknowledging the non-independence of adjacent sensors or time points in an EEG recording. For the clustering procedure, we conducted the ANOVA at each time point and each sensor in the EEG recording. Afterwards, adjacent time points and sensors were clustered together if the p-value of the ANOVA was $p < 0.01$. The clustering procedure was constrained such that each cluster had a minimum length of 10ms and a minimum of 3 sensors. We limited the search to the time period between 100ms and 800ms. We used a stricter p-value threshold than may be necessary, because lower p-values favor smaller clusters, which may be more useful for estimating the temporal onset of effects. Additional *post-hoc* analyses with different p-value thresholds ($p < 0.05$, $p < 0.10$) did not significantly change the pattern of results, and are thus not reported here.

After the clustering procedure was conducted, the *F*-values of the ANOVAs of each cluster were summed to calculate the cluster's size. Then, the condition labels of the stimuli were randomly permuted, and the clustering procedure was conducted another 10,000 times. This produces the estimated null distribution of clusters. The clusters were then ordered by their size. Cluster *p*-value was then equated with position in the resulting null distribution, with the top 5% of clusters corresponding to $p < 0.05$.

The Grammatical vs. Consonant String analysis resulted in several clusters that were significant at the corrected alpha = 0.0167 (Fig. 1E). The largest cluster consisted of 22 centro-parietal sensors, 347–792ms, $p < 0.01$. This cluster showed a sustained positive response for Consonant String trials and a sustained negative response for Grammatical trials. The earliest significant cluster was centered on 14 posterior and right lateral sensors, 220–288ms, $p = 0.0119$. This cluster showed a greater positivity for Grammatical trials compared to Consonant String trials. These clusters suggest a relatively uniform sustained difference in evoked response between Grammatical and Consonant String stimuli; the latter are immediately detectable as non-lexical, and thus none of the subsequent linguistic processes are likely to be engaged, unlike the other 3 conditions.

The Grammatical vs. Scrambled analysis resulted in one significant cluster at the corrected alpha level (Fig. 1E). This cluster consisted of three sensors on the right anterior lateral surface, 320–383ms, $p = 0.0162$. This cluster showed a sustained positive activation for the Grammatical conditions, and a sustained negative response for the Scrambled conditions. Importantly, this does not exhibit the typical N400 waveform morphology or scalp topography, which is a more focal peak spanning from 200–300ms, and usually over centro-parietal and posterior sensors. However, this time window is approximately the same as Wen et al's (2019) N400 finding and approximately 120ms later than Fallon & Pylkkänen's (2023) MEG findings for three word sentences.

The Grammatical vs. Ungrammatical analyses resulted in no significant clusters. This was the case with clustering thresholds of $p < 0.01$, $p < 0.05$, $p < 0.10$, or $p < 0.30$, and with uncorrected alpha = 0.05. Thus, the failure to identify a significant cluster is less likely to be due to the stringentness of our clustering parameters or *p*-value correction.

To determine whether the lack of a difference between Grammatical vs. Ungrammatical conditions indexes a failure to notice the singular vs. plural morphology (i.e., *-s, -Ø* on the –subject noun phrase; *-Ø, -s* on the verb), we conducted a *post hoc* analysis comparing the singular Grammatical trials (*the dog chases a ball*) and plural Grammatical trials (*the dogs chase a ball*). Using the same analysis parameters, we found a significant cluster consisting of 4 central right electrodes, 126–152ms, $p = 0.041$. This cluster can be seen in Fig. 1F. Afterwards, we conducted a *post hoc* Grammatical vs. Ungrammatical analysis nested within the singular trials and the plural trials. We did this, because some behavioral findings demonstrate asymmetrical sensitivity to number errors in subject-verb agreement for plural controllers vs. singular controllers (Wagers et al. 2009). We failed to find any significant clusters distinguishing Grammatical vs. Ungrammatical conditions, nested within singular or plural trials.

3.4 Single-trial Results

A two-stage regression analysis – which allows for the incorporation of multiple continuous predictors – was performed in order to investigate whether and when individual lexical items affected the neural response. In the first stage, within each subject, we fit an ordinary least squares multiple regression to each time point and sensor. This yields, for each subject, a β coefficient for each predictor at each sensor and time point. The first stage was performed using MNE-Python and was performed in the same time window as the previous ANOVA analyses: 100-800ms.

Predictors included: the forward surprisals of words 2, 3, 4 and 5; the *backward* surprisals of words 4, 3, 2, and 1; the cosine similarity of words 2 and 3 (subject noun and verb), and the cosine similarity of words 3 and 5 (verb and object noun). Surprisal (Hale 2001; Levy 2008) is the negative log conditional probability of a word given the preceding context: $-\log P(w_n|w_1,w_2,\ldots,w_{n-1})$. Given *the dogs chase a ball*, the forward surprisal of word 5 is the negative log conditional probability of *ball* given *the dogs chase a* __. By backward surprisal, we mean the negative log conditional probability of a word given the *right-to-left* context (as if 'reading backward,' at least in English, which is read from left to right). Given *the dogs chase a ball*, the backward surprisal of word 1 is the negative log conditional probability of *the* given *ball a chase dogs* __. Cosine similarity is a measure indicating the similarity between two vectors.

Both forward and backward surprisal were estimated using GPT-2 (Radford et al. 2019), using publicly available scripts (Noureddine 2024). The base 2 logarithm was used for the calculation. Forward surprisal is highly correlated with N400 amplitude in serial presentation of sentences (e.g., Frank et al. 2015; Lowder et al. 2018) and taken to index the extra processing needed to access a lexical item in context. To our knowledge, backward surprisal is not commonly studied in studies using serial presentation. The surprisal predictors are investigated because it is possible that expectancy or transition probability – as operationalized by surprisal from the GPT-2 language model – may exert an effect on the recorded EEG signal.

We also included lexical semantic association between the subject noun (*dog*) and the verb (*chase*), and the verb and the object noun (*ball*). This was included as a coarse measure of meaning relation above a single word, or of some thematic interpretation (Lenci 2018). Lexical semantic association was estimated by taking the cosine similarity of the word embeddings from pretrained, 300 dimensional fastText vectors (Bojanowski et al. 2017).

Predictors were standardized before model estimation (i.e., shifted to mean zero and scaled to unit variance). Only Grammatical trials are analyzed: the Consonant String trials are devoid of lexical content; and surprisal is inappropriate in the context of the Scrambled and Ungrammatical trials, which diverge from the distribution of the language model training data (i.e., human-generated text).

In the second stage, we conduct an across-participants analysis in which, for each variable separately, the β values are entered into a two-tailed one-sample *t*-test to determine, at each time point and each sensor, whether their values are significantly different from 0. For estimating significance, we perform a spatio-temporal cluster-based permutation test over the resulting *t*-values. Of note, we use *the same* clustering parameters as those used in the ANOVA spatio-temporal cluster-based permutation analysis: adjacent time points and sensors (of the same polarity) were clustered together if their *p*-value was $p < 0.01$, with the additional constraint that each cluster had a minimum length of 10ms and a minimum of 3 sensors. *T*-values were summed to calculate a cluster-level statistic. A null cluster distribution was estimated by repeating this procedure for 10,000 permutations in which the β value of each participant was randomly shuffled with 0. Cluster *p*-value was equated with position in the null distribution with the bottom 2.5% and top 2.5% being considered statistically significant. The second stage was carried out using Eelbrain.

A significant positive-going cluster was found for the forward surprisal of word 4 consisting of 5 sensors, 238–285ms, $p = 0.047$. A significant positive-going cluster was found for the forward surprisal of word 5 consisting of 3 right sensors, 619–80ms, $p = 0.023$. A significant negative-going cluster was found for the backward surprisal of word 3 consisting of 9 sensors, 392–609ms, $p = 0.007$. A significant negative-going cluster was found for the backward surprisal of word 2 consisting of 3 sensors, 554–801ms, $p = 0.010$. The topographic plots and time series for these clusters can be visualized in Figure 1G. No other significant clusters were found for the predictors of interest.

4. Discussion

Our experiment contributes to the growing literature examining early brain responses, <400ms, to grammatical properties of written sentences displayed rapidly in parallel (Asano & Yokosawa 2011; Snell & Grainger 2017; Wen et al. 2019; Fallon & Pylkkänen 2023; Flower & Pylkkänen 2024; Krogh & Pylkkänen 2024). We found that readers are capable of detecting familiar words vs. non-words ~200ms post-sentence onset, and capable of detecting whether the arrangement of the words maps onto a syntactic representation by ~300ms post-sentence onset. However, we found no evidence that these processes detect subject-verb agreement errors in English, suggesting some limits on the detail of this analysis. Moreover, in an exploratory analysis, we observed that the effects of surprisal of individual lexical items do not surface systematically in the N400 time-window, as overwhelmingly observed in serial presentation. Instead, we observed a non-uniform effect, ranging from ~250ms to ~600ms post-sentence onset.

We suggest that this early, uniform response to grammatical vs. non-grammatical stimuli 150–300ms followed by a non-uniform, inconsistent effect of lexical items is of methodological interest to the cognitive neuroscience of language and psycholinguistics. Whereas experiments in reading of serially-presented sentences are useful for studying the brain's response to *words* in a context, experiments in reading of parallel presented sentences are useful for studying the brain's response to *syntactic structures*. Put differently, for cognitive neuroscientists interested how the brain processes and represents syntactic information, we suggest that comparing and contrasting sentence presented serially with sentences presented in parallel may be a powerful tool for isolating brain activity corresponding to syntactic structure, abstracting away from the impact of lexical material or the memory and attention mechanisms needed to support a representation that evolves over several seconds, in the case of typical serial presentation paradigms.

4.1 Rapid Grammatical Processing and Failure to Notice Agreement Errors

One contribution of our findings to the growing literature on the sentence superiority effect is the differences between scrambled sentences, sentences with agreement errors, and consonant strings. Previous results demonstrate compellingly that sentences are processed more quickly and recalled more accurately than scrambled sentences or non-sentence noun lists. However, it is not immediately evident that this demonstrates any kind of rapid grammatical analysis, since presumably fluent readers have practiced reading sequenuences of characters that approximate sentences more than they have the sequences of characters observed in scrambled sentences (Snell & Grainger 2017; 2019; Wen et al. 2019; Flower & Pylkkänen 2024) or noun lists (Fallon & Pylkkänen 2023). Put differently, from previous results, it's difficult to reject the hypothesis that word shapes and letter *n*-gram frequency may partially drive participants' early neural responses. The finding that Consonant Strings diverge from Grammatical conditions at earlier times and in different topographies than Scrambled sentences, and exhibit more drastic reduction in accuracy in behavioral responses, suggests that comprehenders do engage with these stimuli very differently. Anecdotally, participants found the Consonant String conditions to be incredibly difficult to remember, whereas even the Scrambled sentence conditions were partially understood, as evidenced by the higher matching accuracies. This suggests that even scrambled sentences can engage some linguistic processing compared to non-lexical controls. While there may exist a gradient of familiarity between Consonant String - Scrambled - Grammatical, this also suggests that previous sentence superiority effects cannot be simply reduced to detection of familiar letter patterns vs. unfamiliar letter patterns, since this would predict a more similar response to Consonant String and Scrambled conditions in our study. In other words, we take the sentence superiority effect to index grammatical analysis.

One of our surprising results is that grammatical features may be detectable before the meaning of individual lexical items are accessed and integrated into the interpretation of the sentence. Classical theories in sentence processing have privileged the role that syntactic information plays in the interpretation of sentences ('syntax-first' models; Frazier & Fodor 1978; Friederici 2002). Moreover, theories of morphosyntactic processing have suggested that orthographic and morphosyntactic features may be accessed first before some interpretive processes (Rastle et al. 2000, 2004, 2008; Lewis et al. 2010; Solomyak & Marantz 2011; Fruchter & Marantz 2015). We suggest an analogy may be appropriate for sentences processed using rapid parallel visual presentation. We propose that basic structural features, such as the distribution and position of short functional words or morphemes (*the, a, -s*) is a salient visual cue as to the basic form of the sentence. In other words, short sentences may be read in much the same way that complex words are, by an initial, first-pass rapid analysis of the structure based on orthographic features, followed by a more detailed analysis of the structure and the meaning of its parts. This approach is attractive, because it allows for more closely aligning the reading processes supporting word and sentence reading. This result would be consistent with other findings in psycholinguistics suggesting a blurry boundary between sentences and words (Krauska & Lau 2023, Yu & Lau 2024), and consistent with positions in theoretical linguistics that eschew the distinction between words and sentence constructions (Halle & Marantz 1993).

Although we advocate for a decoupling of the sentence structure from activation of its lexical items, the analogy to morphosyntactic processing is not fully supported by our results. Neural responses to morphologically complex words show a response in left fusiform gyrus ('visual word form area') in the M170 that is sensitive to the stem-to-whole word transition probability, suggesting simultaneous access of both the whole word and its stem, although the corresponding EEG effect may surface somewhat later (Holcomb & Grainger 2006; Grainger & Holcomb 2009; Morris et al. 2013). Our closest proxy to transition probability in this study is the forward surprisal measures in the single-trial analyses. However, these forward surprisal measures do not show anything suggestive of a stem-to-word transition probability effect at the time window expected. Moreover, it's unclear whether the 'stem' in the context of a sentence should be the first word, the verb, the predicate/VP, or something else.

Finally, our single trial analysis only focused on the grammatical trials. This means that our neural responses to the estimates of forward and backward surprisal are all from the same syntactic template. This means we cannot, in the same analysis, test for properties of the grammatical structure and the lexical materials simultaneously. A future study that includes a range of grammatical structures, and a range of frequencies or surprisal values may help elucidate the relationship between rapid structural processing in complex words vs. simple sentences.

If we continue with the suggestion that grammatical structure is accessed for rapidly presented sentences, then we must explain the lack of sensitivity to agreement violations. Like Fallon & Pylkkänen's (2023) and Flower & Pylkkänen's (2024) findings, our results suggest some limits on the kinds of deviations from grammatical structure that result in detectable differences in the neural signal in rapid parallel visual presentation given the current presentation timing scheme. It's unclear whether the failure to observe a deviation in neural signal is attributable to limitations on the visual perceptual system to detect the visual reflexes of these grammatical features, or whether they are 'corrected' in a post-stimulus presentation processing stage after the image offset has disappeared. Importantly, the only visual indicator of agreement for number agreement in English present tense verbs is whether there is an *-s* suffix on the verb or subject noun, which may be easy to miss in parallel reading.

If the matter is simply the salience of the visual stimuli, then research in other languages or other grammatical phenomena may reveal a more striking earlier response in rapid parallel reading. Alternatively, if the matter concerns use of top-down knowledge of syntactic structure to 'correct' a minor deviation from a grammatical structure, perhaps after the visual stimulus has already been removed from the screen, then we might expect these corrections to occur regardless of the relative visual salience of the error. Another attractive alternative is that comprehenders may ignore some morphosyntactic information that is not conducive to identifying the semantic 'gist' of a sentence (Asano & Yosokawa 2011), which may be consistent with 'good enough' parsing models (Ferreira & Patson 2007).

To our knowledge, the robustness of a grammatical error relative to its visual salience is not often considered in sentence processing research. However, Nevins et al. (2007) report on a study of agreement processing in Hindi in which grammatical person mismatch elicits significantly greater P600 responses than other kinds of agreement violations. They suggest that this may be, in part, due to the fact that person agreement mismatches result in a much larger discrepancy between the expected, grammatical orthographic form of the verb and its observed ungrammatical form. If so, this may suggest a fruitful program of future research.

4.2 Timing of Lexical Effects

Classically, the N400 is observed to inversely correlate with the accessibility of a lexical item or other meaningful stimulus (Kutas & Hilyard 1984; Lau et al. 2008; Bornkessel-Schlesewkys & Schlesewsky 2019). The interpretation of the N400 component as corresponding to lexico-semantic processing appears to be key in Wen et al.'s (2019) model, in which lexical activation co-occurs and is facilitated by early syntactic processing. In contrast, our results suggest a distributed effect of different lexical items, both with forward and backward surprisal exhibiting effects for different words at different time points. We propose that this temporal dissociation between the sentence superiority effect and the activation of lexical items indicates that lexical items are not necessarily processed in parallel (White et al. 2019), but rather may sometimes be recovered or inferred from context (Staub 2023) or processed later from an afterimage in visual memory. Future studies manipulating the lexical frequencies and surprisal values of words in different positions may also clarify the degree to which implausible, unexpected, or contextually unusual words are recoverable in parallel reading (Asano & Yokosawa 2011).

We cannot confidently rule out the hypothesis that individual subjects or individual trials include saccades, enabling some amount of serial reading. In other words, our results suggesting different time points for lexical items in different positions may be indicative of multiple 'N400s'. However, we find this unlikely, since our trial structure was designed to make it challenging to read one portion of the sentence, plan and execute a saccade, and finish reading the sentence before the stimulus disappears. Another strategy that participants could have taken is varying the initial fixation point during the task. Some participants reported that they looked purposefully just to the left of the initial fixation cross, so they could maximize the amount of the second half of the sentence that appears in the parafovea, while fixating on the first 1-3 words. In the future, we may mitigate this effect by slightly jittering the horizontal position of the sentence, and using eye-tracking to monitor where participants' eyes are fixated to correlate timing of lexical responses with fixation location.

We also note that the order of the distinct lexical effects appear to be ordered 'inside out;' beginning with sensitivity to lexical surprisal of material in the center of the presumed fixation point or just to its left, ending with sensitivity to lexical surprisal of material at the beginning and end of the sentence. Again, whether this indicates a serial order of operations that occur in parallel reading, or whether this indicates a range of between-item or between-trial fixation strategies, can be better studied by directly manipulating the presentation position of the words and monitoring participants' eye movements using eye-tracking.

Lastly, using a similar single-trial analysis with MEG, Flower & Pylkkänen (2024) observe that the frequency of bigrams in 4-word sentences (*all cats are nice*) exert significant effects ~250ms, and in sentences with transpositions (*all cats nice are*) these effects arise at later times. They suggest this result may indicate an earlier, parallel processing stage for grammatical stimuli, followed by a later, serial processing stage of the ungrammatical stimuli. Although they suggest a different hypothesis, our results are not necessarily inconsistent. First, they measure bigram frequency and transition probability, whereas we use full conditional probability estimated from GPT-2. Full conditional probability as estimated by GPT-2 outperforms n-gram models and LSTM models for reading time data (Hao et al. 2020), and corresponds to the integration of a word into the entire context. By contrast, bigram frequency may index more 'bottom-up' processes, i.e., identification of string units above the word level. Secondly, because their stimuli are shorter than ours, their early parallel effects for grammatical stimuli may correspond to our earlier responses to the 3rd and 4th word ~200–300ms. One possible interpretation of both sets of results is that words encountered in the fovea and their grammatical relations are processed earlier and in parallel, capturing Flower & Pylkkänen's (2024) findings, whereas words further from the fovea are integrated later, detectable in sentences with longer stimuli.

5. Conclusion

Our theories of language comprehension must be as flexible as the range of forms that language can take. In the visual domain, language can be understood slowly or from a quick glance – when editing a journal article vs. checking a text notification. Studies on parallel reading may help us more clearly understand how syntactic information is processed and represented in the brain, distinct from the meaning of the individual words in the structure. Here, we showed further evidence that sentences are a special perceptual unit in the reading of quickly-displayed material, but that rapid psycholinguistic processes may not construct a fully detailed analysis that is capable of noticing formal errors like subject-verb agreement. We also showed that these processes dissociate from the activation of individual lexical items in the sentence. Taken together, this suggests a 'syntax-first' mechanism, in which some grammatical features are accessed even before individual lexical items are, but not morphosyntactic features that are less crucial to identifying the 'gist' of the sentence.

Declaration of Interest
Declarations of interest: none.

References

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., ... & Stegagno, L. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neuroscience letters*, *322*(1), 5-8. https://doi.org/10.1016/S0304-3940(01)02528-9

Antúnez, M., Milligan, S., Hernández-Cabrera, J.A., Barber, H.A., Schotter, E.R. (2022). Semantic parafoveal processing in natural reading: Insight from fixation-related potentials & eye movements. *Psychophysiology* 59(4), e13986. https://doi.org/10.1111/psyp.13986

Asano, M., Yokosawa, K. (2011). Rapid extraction of gist from visual text and its influence on word recognition. *The Journal of General Psychology* 138, 127–154. https://doi.org/10.1080/00221309.2010.542510

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, *5*, 135-146. https://doi.org/10.1162/tacl_a_00051

Bornkessel-Schlesewsky, I., Schlesewsky, M. Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology* 10, 298. https://doi.org/10.3389/fpsyg.2019.00298

Brasoveanu, A., & Dotlačil, J. (2020). *Computational cognitive modeling and linguistic theory*. Springer Nature. https://library.oapen.org/handle/20.500.12657/39529

Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2023). Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions. *Elife 12*, e85012. https://doi.org/10.7554/eLife.85012

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and cognitive processes*, *13*(1), 21-58. https://doi.org/10.1080/016909698386582

Fallon, J. & Pylkkanen, L. (2023). MEG evidence for ultrarapid phrase structure detection at 100-200ms in parallel visual presentation. Poster presented at the *15th Annual Meeting of the Society for the Neurobiology of Language*, Marseille, France.

Ferreira, F., Patson, N. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Flower, N. & Pylkkanen, L. (2024). The spatiotemporal dynamics of bottom-up and top-down processing during at-a-glance reading. *bioRxiv*. https://doi.org/10.1101/2024.02.26.582140

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5(4), 519–559. https://doi.org/10.1007/BF00138988

Frazier, L., Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition* 6(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *Neuroimage*, *20*, S8-S17. https://doi.org/10.1016/j.neuroimage.2003.09.003

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain research*, *1146*, 50-58. https://doi.org/10.1016/j.brainres.2006.08.038

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological reviews*, *91*(4), 1357-1392. https://doi.org/10.1152/physrev.00006.2011

Fruchter, J., Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language* 143, 81–96. https://doi.org/10.1016/j.bandl.2015.03.001

Grainger, J., Holcomb, P.J. (2009). An ERP investigation of orthographic priming with relative-position and absolute-position primes. *Brain Research* 1270; 45–53. https://doi.org/10.1016/j.brainres.2009.02.080

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 267. https://doi.org/10.3389/fnins.2013.00267

Gwilliams, L., Marantz, A. (2018). Morphological representations are extrapolated from morpho-syntactic rules. *Neuropsychologia* 114, 77–87. https://doi.org/10.1016/j.neuropsychologia.2018.04.015

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the North American Chapter of the Association for Computational Linguistics* 2, 1–8. https://doi.org/10.3115/1073336.1073357

Halle, M., Marantz, A. (1993). Distributed morphology and pieces of the lexicon. In M. Halle & S.J. Keyser (eds.), *The View from Building 20*. MIT Press.

Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics,* 75–86. https://doi.org/10.18653/v1/2020.cmcl-1.10

Holcomb, P.J., Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience* 18(10), 1631–1643. https://doi.org/10.1162/jocn.2006.18.10.1631

Kaan, E. (2007). Event‐related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571-591. https://doi.org/10.1111/j.1749-818X.2007.00037.x

Krauska, A., Lau, E. (2023). Moving away from lexicalism in psycho- and neurolinguistics. *Frontiers in Language Science* 2, 1125127. https://doi.org/10.3389/flang.2023.1125127

Krogh, S. & Pylkkanen, L. (2024). Manipulating syntax without taxing working memory: MEG correlates of syntactic dependencies in a Verb-Second language. *bioRxiv*. https://doi.org/10.1101/2024.02.20.581245

Kutas, M., Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. https://doi.org/10.1038/307161a0

Lau, E., Phillips, C., Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews: Neuroscience* 9(12), 920–933. https://doi.org/10.1038/nrn2532

Lenci, A. (2018). Distributional models of word meaning. *Annual review of Linguistics*, *4*, 151-171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewis, G., Solomyak, O., Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain & Language* 118(3), 118–127. https://doi.org/10.1016/j.bandl.2011.04.004

Lewis, R., Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lowder, M.W., Choi, W., Ferreira, F., Henderson, J.M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science* 42(S4), 1166–1183. https://doi.org/10.1111/cogs.12597

Maris, E., Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Massol, S., Mirault, J., Grainger, J. (2021). The contribution of semantics to the sentence superiority effect. *Scientific Reports* 11, 20148. https://doi.org/10.1038/s41598-021-99565-6

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *cortex*, *47*(8), 908-930. https://doi.org/10.1016/j.cortex.2011.02.019

Morris, J., Grainger, J., Holcomb, P.J. Tracking the consequences of morpho-orthographic decomposition of words using ERPs. *Brain Research* 1529, 92–104. https://doi.org/10.1016/j.brainres.2013.07.016

Nevins, A., Dillon, B., Malhotra, S., Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research* 1164, 81–94. https://doi.org/10.1016/j.brainres.2007.05.058

Noureddine, S. (2024). GPT-2 for Psycholinguistic Applications. GitHub repository. https://github.com/samer-noureddine/GPT-2-for-Psycholinguistic-Applications

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739-773. https://doi.org/10.1006/jmla.1995.1033

Pegado, F., & Grainger, J. (2020). A transposed-word effect in same-different judgments to sequences of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1364. https://doi.org/10.1037/xlm0000776

Pegado, F., Wen, Y., Mirault, J., Dufau, S., & Grainger, J. (2021). An ERP investigation of transposed-word effects in same-different matching. *Neuropsychologia*, *153*, 107753. https://doi.org/10.1016/j.neuropsychologia.2021.107753

Pylkkänen, L., Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences* 7(5), 187–189. https://doi.org/10.1016/S1364-6613(03)00092-5

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9. https://openai.com/research/better-language-models

Rastle, K., Davis, M.H., Marslen-Wilson W.D. (2000). Morphological and semantic effects in visual word form recognition: A time-course study. *Language and Cognitive Processes* 15(4/5), 507–537. https://doi.org/10.1080/01690960050119689

Rastle, K., Davis, M.H., New, B. (2004). The broth in my brother's brothel: Morpho-orthoraphic segmentation in visual word recognition. *Psychonomic Bulletin & Review* 11(6), 1090–1098. https://doi.org/10.3758/BF03196742

Rastle, K., Davis, M.H. (2008). Morphological decomposition on the basis of orthography. *Language and Cognitive Processes* 23(7–8), 942-971. https://doi.org/10.1080/01690960802069730

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372. https://doi.org/10.1037/0033-2909.124.3.372.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*, 5-35. https://doi.org/10.3758/s13414-011-0219-2

Snell, J., Grainger, J. (2017). The sentence superiority effect revisited. *Cognition* 168, 217–221. https://doi.org/10.1016/j.cognition.2017.07.003

Snell, J., Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537-546. https://doi.org/10.1016/j.tics.2019.04.006

Solomyak, O., Marantz, A. (2011). Evidence for early morphological decomposition in visual word recognition: A single-trial correlational study. *Journal of Cognitive Neuroscience* 22(9), 2042–2057. https://doi.org/10.1162/jocn.2009.21296

Staub, A. (2023). A commentary on the 'sentence superiority effect': Almost certainly guessing, not parallel interactive activation. *PsyArXiv Preprints.* https://doi.org/10.31234/osf.io/hzaqr

Taft, M., Forster, K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14, 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of memory and language*, *61*(2), 206-237. https://doi.org/10.1016/j.jml.2009.04.002

Wen, Y., Snell, J., Grainger, J. (2019). Parallel, cascaded, interacting processing of words during sentence reading. *Cognition* 189, 221–226. https://doi.org/10.1016/j.cognition.2019.04.013

White, A., Boynton, G.M., Yeatman, J.D. (2019). You can't recognize two words simultaneously. *Trends in Cognitive Science*s 23(10), 812–814. https://doi.org/10.1016/j.tics.2019.07.001

Yu, X., Lau, E. (2024). Electrophysiological responses to syntactic and 'morphological' structures: Evidence from Mandarin Chinese. *bioRxiv* 2024.01.31.578104 https://doi.org/10.1101/2024.01.31.578104

Zweig, E., Pylkkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes* 24(3), 412–439. https://doi.org/10.1080/01690960802180420

Appendix A.

(Stimuli)

1	the girls buy a book
2	book girls the buy a
3	the girls buys a book
4	bxb tgffb twv q xplh
5	some boys eat an apple
6	an boys some eat apple
7	some boys eats an apple
8	tpbt tjzg wxs gd fhdvm
9	the dogs chase a ball
10	a dogs chase ball the
11	the dogs chases a ball
12	thj rjxb zkhtb w lhct
13	the cops find a clue
14	clue cops find a the
15	the cops finds a clue
16	vrl kgmd mbdt b rdhk
17	the cats see a rat
18	see a cats the rat
19	the cats sees a rat
20	mfr vfhs bqc r vjz
21	the guys make a coat
22	make guys a the coat
23	the guys makes a coat
24	snj vhvs dvkr l phvc
25	the maids clean a room
26	a maids the room clean

27	the maids cleans a room
28	dln knnjd cdwvs l plbj
29	the kings write a rule
30	rule kings the a write
31	the kings writes a rule
32	zxp vnccj jprjz s cfcs
33	the queens read a note
34	the note read queens a
35	the queens reads a note
36	zvj qmjcrl hghw p bsfm
37	the dads take a bath
38	dads bath take the a
39	the dads takes a bath
40	pgh pvgr nsth j ztnm
41	the teams win a game
42	win the teams a game
43	the teams wins a game
44	hhq pqjfn mrb d cmss
45	the friends sell a car
46	sells the a friends car
47	the friends sells a car
48	vrg vslzhvw bbnbm r phl
49	the moms drink a beer
50	drink moms the a beer
51	the moms drinks a beer
52	jct hnnx hpbzp k mqqr
53	the chicks like a song
54	song chicks a the like
55	the chicks likes a song

56	ddp lvjppz mnwv p rxts
57	the nurses heal a wound
58	a the heal wound nurses
59	the nurses heals a wound
60	tcm ckhvcs gvgm k xnfpq
61	the kids taste a treat
62	kids treat taste a the
63	the kids tastes a treat
64	szl xqgk dmwgd b gpvqj
65	the tykes ride a bike
66	ride a the tykes bike
67	the tykes rides a bike
68	jwq nsjjr cwlv n ppqc
69	the teens draw a shape
70	the teens a draw shape
71	the teens draws a shape
72	csh nlflq mmpw h fnffp
73	the dudes paint a house
74	paint dudes the house a
75	the dudes paints a house
76	wlv zsjtd mrtss c mmfnr
77	the friends know a joke
78	a friends know joke the
79	the friends knows a joke
80	rtk nbwhvvg ldjq k rnlt
81	the grooms ride a horse
82	grooms the ride horse a
83	the grooms rides a horse
84	tzc gbqcvg vcjs q qxrnj

85	the brides give a speech
86	a speech give brides the
87	the brides gives a speech
88	wpw hrgljr vgdg r dptpkt
89	the knights kill a bear
90	the a kill knights bear
91	the knights kills a bear
92	qhs znjhzgg qlbm c pkqm
93	the tots sell a toy
94	a the sell tots toy
95	the tots sells a toy
96	rnb tssn qcqt z ndg
97	the gang steals a bike
98	steals gang bike the a
99	the gang steal a bike
100	mnc hnss slvldl r ttjk
101	the squad sees a movie
102	a squad sees movie the
103	the squad see a movie
104	tls qmnjr mntx f nmjbv
105	the sleuth spots a clue
106	clue sleuth spots a the
107	the sleuth spot a clue
108	tsj hnpkgk krgvw c gpdj
109	the dad sips a beer
110	beer a sips dad the
111	the dad sip a beer
112	ngv vwm vqzj t gkxq
113	the mom sews a patch

114	sews a the mom patch
115	the mom sew a patch
116	tkw gjc fxwv m sfsgm
117	the kid sees a star
118	sees kid the a star
119	the kid see a star
120	fcl wcz qkkf f ztvb
121	the teen writes a note
122	writes teen the a note
123	the teen write a note
124	ngz pmqh qcppxd d bqrx
125	the groom wears a suit
126	a groom suit wears the
127	the groom wear a suit
128	vbm frzwk mtpgn r frpx
129	the priest reads a verse
130	the verse reads priest a
131	the priest read a verse
132	qqn ltbfnk lmkvj v bgpzs
133	the nurse takes a test
134	nurse test takes a the
135	the nurse take a test
136	mvp kjsxd hxbgr h nhfz
137	the cop drives a car
138	drives a cop the car
139	the cop drive a car
140	fmb chp kxvqsq d xtl
141	the girl hums a tune
142	girl hums a the tune

143	the girl hum a tune
144	qpf gsgz tllg m rkzp
145	the nurse chews a snack
146	chews nurse a the snack
147	the nurse chew a snack
148	wxh gkmzz ttxcd z fngjm
149	the boy likes a game
150	likes boy game a the
151	the boy like a game
152	vcd bgz lqzvz r sgll
153	the wife climbs a tree
154	tree wife climbs a the
155	the wife climb a tree
156	tkm ffxq nlcxsm n lfhm
157	the king wears a hat
158	king the wears hat a
159	the king wear a hat
160	ddf bmss lmwww z vqn
161	the chick gets a gift
162	a the chick gets gift
163	the chick get a gift
164	bgv gmxsv kdbz l mzxf
165	the guy kills a bee
166	guy kills a the bee
167	the guy kill a bee
168	blz zpj vdrxf h kdk
169	the tot tears a page
170	page tot tears a the
171	the tot tear a page

172	zjn tpx mjstb v wlcn
173	the maid lifts a chair
174	a maid chair lifts the
175	the maid lift a chair
176	bcm rwwm gkmqp p mvmjp
177	the aunt buys a doll
178	aunt doll buys a the
179	the aunt buy a doll
180	zst cfdg kfhb r tbcw
181	the cat smacks a ball
182	ball a smacks the cat
183	the cat smack a ball
184	tpz znw djzpbv w dflq
185	the dude shoots a gun
186	the shoots a dude gun
187	the dude shoot a gun
188	xzj vxtc hzzgnx m xbw
189	the duke swings a sword
190	swings duke the a sword
191	the duke swing a sword
192	lst sbrd lpqjlk k rlsjw