Parafoveal semantic integration in natural reading

Yali Pan[1], Steven Frisson[1], Kara D. Federmeier[2], Ole Jensen[1]*

[1]Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK
[2]Department of Psychology, Program in Neuroscience, and the Beckman Institute for Advanced Science and Technology, University of Illinois, Champaign, Illinois, USA

*Corresponding author. Email: O.Jensen@bham.ac.uk

Abstract

While humans have a remarkable ability to read and comprehend text, the neuronal mechanisms supporting natural reading remain elusive. It is debated whether upcoming words in the parafovea are processed semantically. Here we show that the brain can indeed detect that a word is contextually implausible even before fixating on it. We demonstrated this by imperceivably flickering target words at 60 Hz that either were congruent or incongruent with the sentence context. While participants read these sentences, we used MEG to detect the flicker-response prior to saccades to the target words. We found a smaller tagging response for words that were incongruent with the sentence context; this reduction was related to the individual reading speed. We conclude that parafoveal semantic integration occurs during natural reading.

One Sentence Summary

During natural reading, upcoming words are previewed and integrated into the context even before we move our eyes to it.

Main Text

How quickly do we notice that something does not make sense? During natural reading, fixations only last for 225-250 ms, meaning that there is a very short time available to extract a word's meaning, integrate it within the context, and plan the next saccade (*1*, *2*). Interestingly, the next saccade does not always fall onto the next word, as we skip words during reading (*3*). This does not, however, seem to impede comprehension, suggesting that sufficient information about an upcoming (parafoveal) word can at least sometimes be extracted before the saccade. What has remained unresolved is whether processing the parafoveal word at a semantic level is an exception, done only when the next word is short, high in frequency and highly predictable, or whether it is the standard way of reading. Moreover, even if semantic information can be extracted from the parafovea, it is unknown whether the previewed information is also immediately integrated into the evolving context, thus interacting with the ongoing reading process. As we will describe, our data show that the brain can indeed detect that a word is contextually implausible even before fixating on it, within ~100 ms of fixating on the word *before* the incongruent word.

Studies measuring eye movement patterns have found evidence for the extraction of parafoveal semantic information in Chinese (*4*) and German (*5*), but not in English (*6–9*) unless the next word appears in a highly constraining context (*10*) or is made visually more salient by capitalizing the initial letter (*11*). This mixed evidence makes the extent of semantic parafoveal processing an ongoing question in the context of English reading (*12*). Electrophysiological studies have helped to fill this gap by providing key insights into how

semantics unfolds during language comprehension, but little work has been done with natural reading (although see (*13–18*)). In passive reading paradigms, wherein words are presented foveally one-by-one, often for a fixed amount of time, context-based facilitation of semantic processing can be observed as reductions in the amplitude of the N400 component (*19*), a negative-going ERP response peaking around 400 ms post-word-onset that has been linked to semantic access (*20*, *21*). N400 context effects – smaller responses to words that are more expected (e.g., to "butter" compared to "socks" after "He spread the warm bread with …") – can also be observed for parafoveally-presented words in English, even in the absence of strong constraint or high salience (*13–18*). However, the passive reading paradigm, which prohibits saccading, might not reproduce naturalistic reading. For example, in recent work, we showed that saccades are not trivial artefacts but can modulate the ongoing brain activity and help with word processing in reading (*22*). Thus, to understand normal parafoveal semantic processing during reading, it is critical to be able to measure it under naturalistic conditions.

Here, we aimed to provide direct neural evidence that semantics can be extracted from the parafovea during natural reading and to demonstrate how previewed semantics interacts with the ongoing reading process. Rapid invisible frequency tagging (RIFT) is a novel tool that we have developed to measure neuronal excitability related to the processing of visual flicker (*23–26*), which has been shown to successfully capture parafoveal processing in natural reading (*27*). Here, we applied RIFT to a silent reading task, where the target word was tagged with an underlying patch. The luminance of the patch flickered in a sinusoid of 60 Hz, which was not perceived differently from the background, making it invisible (Fig. 1A). To focus on semantic parafoveal processing while avoiding confounds of word-level predictability, each sentence contained an unpredictable target word that was either congruent or incongruent with the sentence context (see Supplementary Materials for the pre-tests of target predictability and sentence plausibility). Participants (n = 34) read one-line sentences silently while eye movements and brain activity were recorded with an eye tracker and MEG simultaneously. We use the RIFT responses during the fixation of the pre-target word as an indicator of the parafoveal processing of the target word.

Like prior work in English reading (*6–9*), we found no evidence for parafoveal semantic processing in the eye movement data (Fig. 1B, left). A paired t-test comparing first fixation durations on the pre-target word showed no effect of semantic (in)congruency ($t_{(33)} = .84$, $p = .407$, $d = 0.14$, two-sided). However, first fixation durations on the target word were significantly longer when they were incongruent (versus congruent) with the context ($t_{(33)} = 5.99$, $p = 9.83\times10^{-7}$, $d = 1.03$, two-sided pairwise *t*-test; Fig. 1B, right). In addition, we found that the contextual congruity of target words affected later eye movement measures (total gaze duration and the likelihood of refixation after the first pass reading), with additional processing evident when the target words were incongruent with the context compared with congruent (fig. S1).

Next, we estimated the neural responses to the flickering target words in the parafovea by calculating the coherence between the MEG sensors and the tagging signal. A MEG sensor was selected if it showed significantly stronger 60 Hz coherence during the pre-target intervals compared to the baseline intervals (no flicker). After applying a cluster-based permutation test, we found sensor clusters that showed robust tagging responses ($p_{\text{cluster}} < .01$; Fig. 2A), whose sources were localized to the left visual association cortex (Brodmann area 18; Fig. 2B). We then characterized the dynamics of attentional allocation to the flickering target word by calculating the 60 Hz coherence during the fixations of several words around the target word (Fig. 2C). This RIFT response curve shows that a significant amount of

attention was allocated to the target word even three words before, a span of 15.3 ± 2.7 letters long including the spaces between words (M ± SD; see Table S1 for characteristics of words). This is consistent with previous measures of the perceptual span (14-15 letters; (*1*)) based on the moving window technique and indicates that RIFT did not affect attention distribution during natural reading.

Importantly, evidence for semantic parafoveal processing was found using RIFT (Fig. 3). Our key finding was that magnitude of the pre-target coherence was significantly weaker when the following target word was contextually incongruent compared with congruent with the sentence context ($t_{(28)}$ = -2.561, p = .016, two-sided pairwise t-test; Fig. 3B). Next, we conducted a latency estimation and found that the onset of the averaged pre-target coherence was significantly later when previewing an incongruent (116.0 ± 1.9 ms, M ± SD) compared to a congruent target word (91.4 ± 2.1 ms, M ± SD; $t_{(28)}$ = -2.172, p = .039, two-sided; Jack-knife approach; Fig. 3C). We conducted the same analysis for coherence for fixations on the target words and found no significant modulations reflecting the congruity of the target word for either the magnitude or onset latency of the RIFT signal (fig. S2).

We also conducted a correlation analysis to investigate how parafoveal semantic processing related to reading speed (Fig. 4). Here, reading speed was quantified as the number of words read per second from the experimental congruent sentences. We found a positive correlation between the pre-target coherence difference and individual reading speed ($r_{(27)}$ = .503, p = .006; Spearman's correlation). This suggests that fast readers allocate relatively more attention in the attempt to integrate parafoveal words in the sentence context even when they were incongruent.

Why was the pre-target coherence found to be weaker in magnitude (Fig. 3B) and later in onset latency (Fig. 3C) when followed by an incongruent compared with a congruent target word? There are two possible explanations. First, when the parafoveal target word is incongruent, more effort is required to try to integrate the previewed semantics into the prior context. Thus, attention may be allocated more to this internal integration process rather than to the external visual processing of the flickering word that is measured by RIFT. One study has shown that responses to a visual flicker decreased with more attention being directed internally (*28*). Second, our attention continuously shifts from left to right when reading English. However, if the previewed semantic information cannot easily be integrated into the context, fluent reading is interrupted and the allocation of attention to the upcoming word is delayed and/or reduced.

Here, we use RIFT to index the amount of (covert) external attention allocated to the target word. That attentional focus then likely spans different aspects of word processing over time and word position. For instance, a significant amount of attention allocated to the target word was detected even three words in advance (Fig. 2C). Given that this initial increase in RIFT occurred when the target just entered the perceptual span (*1*), it is likely that it coincides with the initial extraction of lower-level perceptual information about the target. The emerging sensitivity of the RIFT signal to target plausibility by about 100 ms into the fixation on the pre-target word suggests that it is by this point in processing that readers had accumulated enough semantic information about the target words and done enough integration of that information with the unfolding sentence context to detect the contextual incongruity. Thus, it seems likely that initial processing of word semantics began even before the pre-target fixation and was spread out across fixations to multiple words.

The rapid parafoveal semantic integration was only captured by RIFT (Fig. 3B and Fig. 3C) but not in the early eye movement measures (Fig. 1B). This is because first fixation durations and the RIFT signal capture different aspects of natural reading. Fixation duration measures

how long people spend fixating on a foveal word, which indicates aspects of overt attention to that word, while RIFT measures covert attention. This was further evidenced by the null result of RIFT when fixating on the target words. Due to more motion-sensitive rod cells in the parafoveal than foveal area, RIFT is possibly more sensitive to the flickering word in the parafovea. Thus, when fixating on the target word foveally, the effect of semantic violation was not found with the RIFT approach (fig. S2) but is observed in the first fixation duration patterns (Fig. 1B). Furthermore, although readers have accumulated enough information from the parafovea so that the attentional allocation is different for the context (in)congruity condition, this process is still too slow to change the saccade planning. Thus, the first fixation duration of the pre-target word was not modulated by the contextual congruity of the target word.

Our data did not show a significant effect of target congruity on the fixation-related fields (FRFs) of the pre-target words (fig. S3). Why did we not observe mN400 effects – either parafoveally or foveally – in our FRFs? It is important to note that all of words we used were of low cloze probability and were embedded in low constraint contexts (see Stimuli in Supplementary material for more details). Previous work has established that the N400 is highly sensitive to expectancy (as indexed by cloze probability) but generally does not differ across words matched in cloze probability as a function of factors such as constraint or plausibility; our findings here thus replicate patterns observed for passive reading (*29*, *30*). Indeed, the lack of an mN400 effect provides supporting evidence that our RIFT effects index semantic processing beyond basic word expectancy – driven, instead, by differences in the ease with which the target words can be integrated with the larger context.

In sum, we applied the RIFT to a natural reading study and showed that attention is already allocated to a given word prior to fixating on it. In distributing attention across text in this manner, readers are not only able to extract semantic information from the target word while it is in parafoveal vision but are also able to integrate the semantics of the target word with the unfolding representation of the sentence. Semantic plausibility can thereby be determined before the target is even fixated – as fast as 100 ms after fixating the pre-target word – and this congruency information is used to further shape attentional allocation. These findings suggest that semantic processing during natural reading unfolds rapidly over time and distributes broadly over words.

References

1. K. Rayner, *Q. J. Exp. Psychol.* **62**, 1457–1506 (2009).
2. O. Jensen, Y. Pan, S. Frisson, L. Wang, *Trends Cogn. Sci.* (2021).
3. K. Rayner, T. J. Slattery, D. Drieghe, S. P. Liversedge, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 514–528 (2011).
4. M. Yan, E. M. Richter, H. Shu, R. Kliegl, *Psychon. Bull. Rev.* **16**, 561–566 (2009).
5. S. Hohenstein, R. Klieg, *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 166–190 (2014).
6. A. W. Inhoff, K. Rayner, *Percept. Psychophys.* **27**, 457–464 (1980).
7. A. W. Inhoff, *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 137–145 (1982).
8. K. Rayner, E. R. Schotter, D. Drieghe, *Psychon. Bull. Rev.* **21**, 1067–1072 (2014).
9. K. Rayner, D. A. Balota, A. Pollatsek, *Artic. Can. J. Psychol. Rev. Can. Psychol.* **40**, 473–483 (1986).

10. E. R. Schotter, M. Lee, M. Reiderman, K. Rayner, *J. Mem. Lang.* **83**, 118–139 (2015).

11. K. Rayner, E. R. Schotter, *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1617–1628 (2014).

12. E. R. Schotter, B. Angele, K. Rayner, *Attention, Perception, Psychophys.* **74**, 5–35 (2012).

13. H. A. Barber, N. Doñamayor, M. Kutas, T. Münte, *Neurosci. Lett.* **479**, 152–156 (2010).

14. H. A. Barber, M. van der Meij, M. Kutas, *Psychophysiology*. **50**, 48–59 (2013).

15. P. J. López-Peréz, J. Dampuré, J. A. Hernández-Cabrera, H. A. Barber, *Brain Lang.* **162**, 29–34 (2016).

16. M. C. Stites, B. R. Payne, K. D. Federmeier, *Cogn. Affect. Behav. Neurosci.* **17**, 475–490 (2017).

17. B. R. Payne, M. C. Stites, K. D. Federmeier, *Psychophysiology*. **56**, 1–15 (2019).

18. B. R. Payne, K. D. Federmeier, *Neuropsychologia*. **106**, 358–370 (2017).

19. M. Kutas, S. A. Hillyard, *Science (80-.).* **207**, 203–205 (1980).

20. M. Kutas, K. D. Federmeier, *Annu. Rev. Psychol.* **62**, 621–647 (2011).

21. K. D. Federmeier, *Psychophysiology*. **59**, e13940 (2022).

22. Y. Pan, T. Popov, S. Frisson, O. Jensen, *bioRxiv* (2022).

23. A. Zhigalov, J. D. Herring, J. Herpers, T. O. Bergmann, O. Jensen, *Neuroimage.* **195**, 59–66 (2019).

24. K. Duecker, T. P. Gutteling, C. S. Herrmann, O. Jensen, *J. Neurosci.* **41**, 6684–6698 (2021).

25. T. P. Gutteling, L. Sillekens, N. Lavie, O. Jensen, *Prog. Neurobiol.* **214**, 102285 (2022).

26. L. Drijvers, O. Jensen, E. Spaak, *Hum. Brain Mapp.* **42**, 1138–1152 (2021).

27. Y. Pan, S. Frisson, O. Jensen, *Nat. Commun. 2021 121*. **12**, 1–9 (2021).

28. L. Kritzman *et al.*, *Neuroimage*. **254**, 119133 (2022).

29. M. Kutas, K. D. Federmeier, *Trends Cogn. Sci.* **4**, 463–470 (2000).

30. M. Kutas, S. A. Hillyard, *Nature*. **307**, 161–163 (1984).

Acknowledgments: The computations described in this paper were performed using the University of Birmingham's BlueBEAR HPC service, which provides a High Performance Computing service to the University's research community.
See http://www.birmingham.ac.uk/bear for more details.

Funding:

James S. McDonnell Foundation Understanding Human Cognition Collaborative Award 220020448 (OJ)

Wellcome Trust Investigator Award in Science 207550 (OJ)

BBSRC grant BB/R018723/1 (OJ)

Royal Society Wolfson Research Merit Award (OJ)

Author contributions:

Conceptualization: YP, OJ

Methodology: YP, OJ

Investigation: YP

Visualization: YP

Supervision: OJ, KF, SF

Writing—original draft: YP, OJ

Writing—review & editing: YP, OJ, KF, SF

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: We have deposited the following data in the current study on figshare (https://figshare.com/projects/149801): the epoch data after pre-processing, the raw EyeLink files, the Psychotoolbox data, and the head models after the co-registration of T1 images with the MEG data. The raw T1 images are not shared due to sensitive personal information (faces). Anonymizing T1 images will remove the informative facial landmarks and make it difficult to construct head models. Therefore, we share the head models instead of the T1 images. Any additional information will be available from the authors upon reasonable request. The experiment presentation scripts (Psychtoolbox), statistics scripts (R), scripts and data to generate all figures (Matlab) are available on GitHub (https://github.com/yalipan666/Semantic)

Figures and tables in the main text

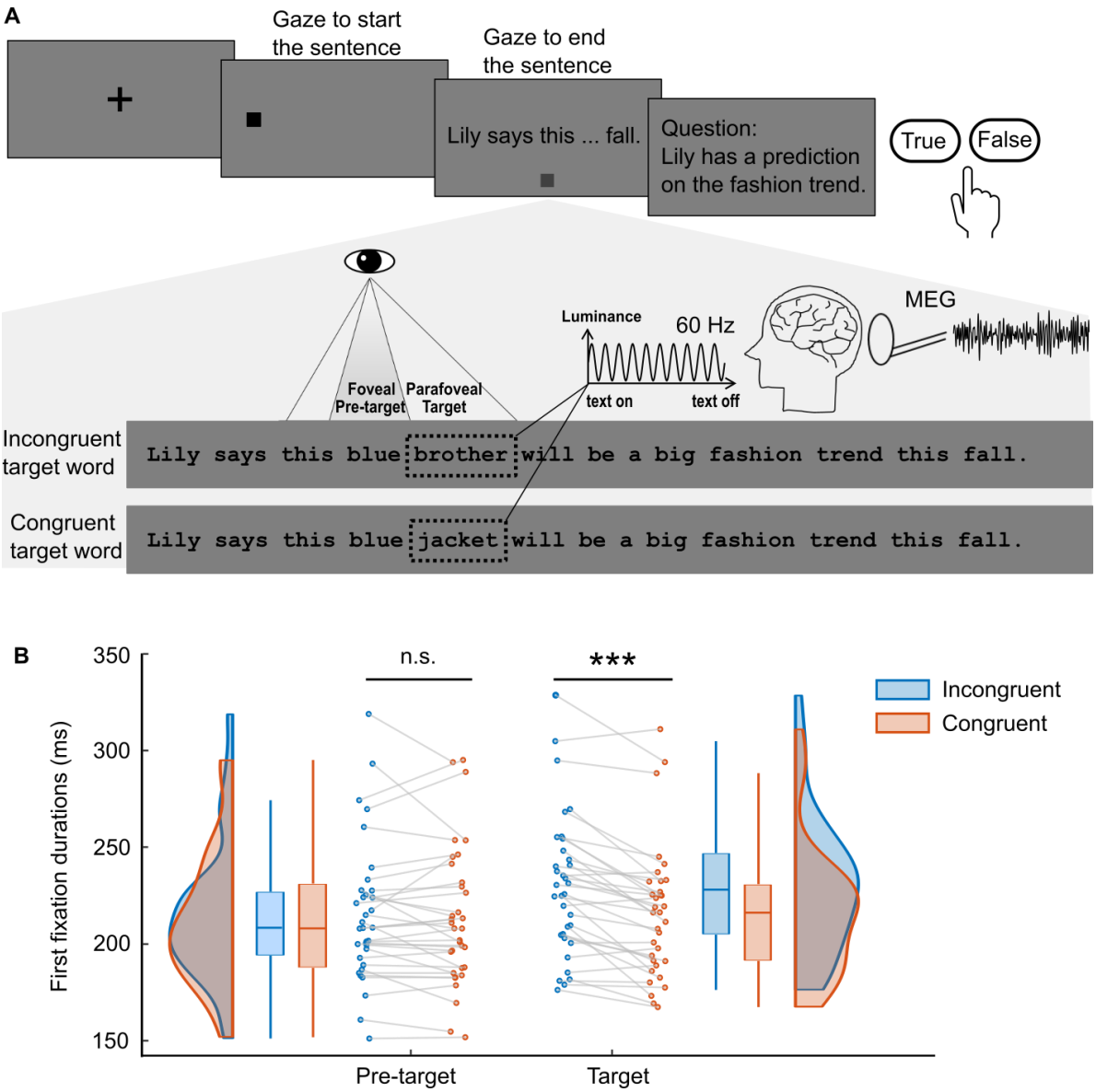

Fig. 1. The paradigm and the eye movement metrics. (**A**) After the presentation of a cross-fixation at the screen centre for 1.2 −1.6 s, a gaze-contingent box appeared near the left edge of the screen. A gaze at the box for 0.2 s triggered the full sentence presentation. Participants (n = 34) read 160 one-line sentences silently while brain activity and eye movements were recorded. Each sentence was embedded with one congruent or incongruent target word (see the dashed rectangle; not shown in the experiment). The target words could not be predicted by the sentence context and were balanced across participants. We tagged the target words by changing the luminance of the underlying patch in a 60 Hz sinusoid throughout the sentence presentation. After reading, gazing at the bottom box for 0.1 s triggered the sentence offset. Twelve percentage of the sentences were followed by a simple yes-or-no comprehension question. (**B**) The first fixation durations on the pre-target and target words when the target words were incongruent (in blue) or congruent (in orange) with the sentence context. Each dot indicates one participant. ***$p < .001$; n.s., not statistically significant.

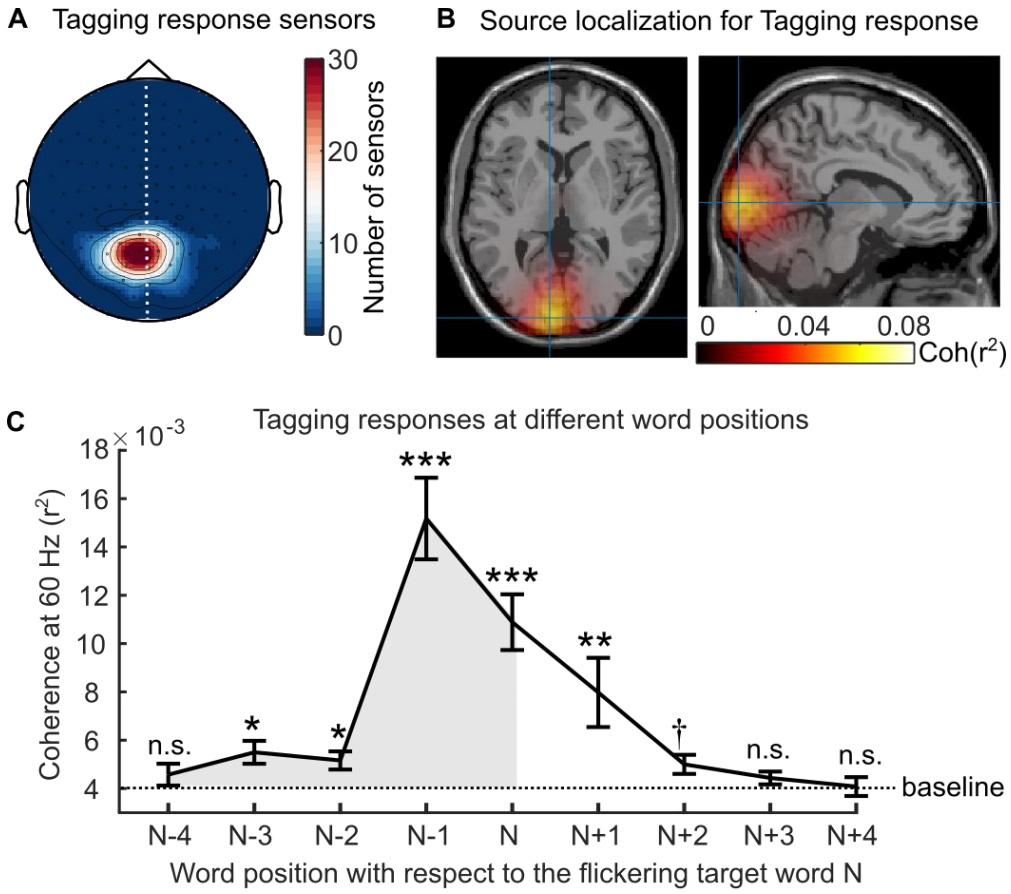

Fig. 2. Neural responses to the rapid invisible frequency tagging (RIFT). (**A**) Topography of the RIFT response sensors over all participants (7.9 ± 4.5 sensors per participant, M ± SD). These sensors showed significantly stronger coherence to the tagging signal during the pre-target intervals (with target words flickering in the parafovea) compared with the baseline intervals (no flicker). Further analyses only included participants who had a RIFT response (n = 29). (**B**) The source of the RIFT response sensors was localized to the left visual association cortex (MNI coordinates [-9 -97 3] mm, Brodmann area 18). (**C**) The averaged 60 Hz coherence over the RIFT response sensors when participants fixated on words at different positions, where N indicated the target words and N-1 indicated the pre-target words. Error bars indicate the SE over participants (n = 29). The shaded area indicates the RIFT responses when pre-processing the flickering target words. We compared the RIFT response at each word position and the baseline (the dashed line). ***$p < .001$, **$p < .01$, *$p < .05$, † $p = .051$; n.s., not statistically significant.

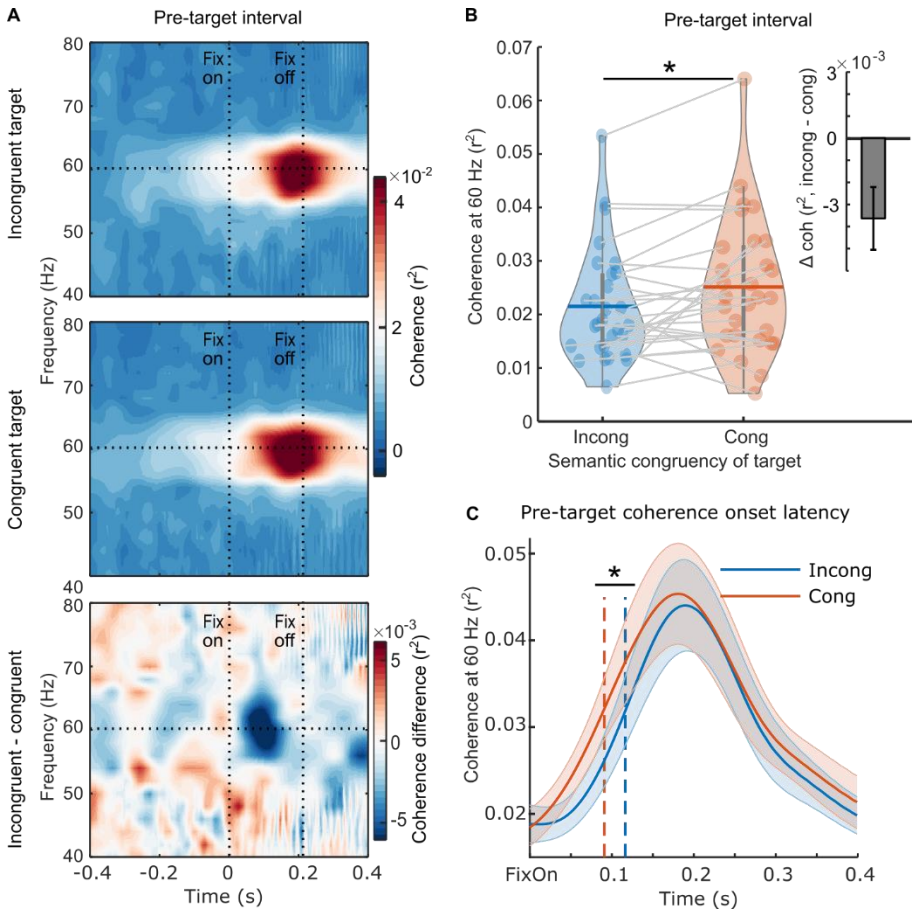

Fig. 3. Neural evidence for parafoveal semantic integration. (**A**) The pre-target coherence spectrum averaged over the RIFT response sensors at the group level (n = 29) when the following target words were incongruent with the sentence context (top panel), congruent with the sentence context (middle panel), or the difference between two conditions (bottom panel). The horizontal line indicated the tagging frequency at 60 Hz. The two vertical lines indicated the first fixation onset of the pre-target words and the averaged fixation offset. (**B**) The averaged 60 Hz coherence during the pre-target intervals with respect to the incongruent and congruent target words. To avoid any contamination of the parafoveal measure with activity from target fixation, we found the minimum pre-target fixation duration across both conditions for each participant (97.4 ± 14.1 ms, M ± SD). Each dot indicates one participant, the horizontal lines inside of the violins indicate the mean values. The upright inserted figure showed the pre-target coherence difference over participants with the error bar as SE. (**C**) The onset latency of the pre-target coherence at the group level (n = 29). The onset latency refers to the time when the coherence curve reached its half maximum, denoted by the dotted lines. Zero time-point indicates the first fixation onset of the pre-target words. The shaded area shows SE around the mean value. *p < .05.

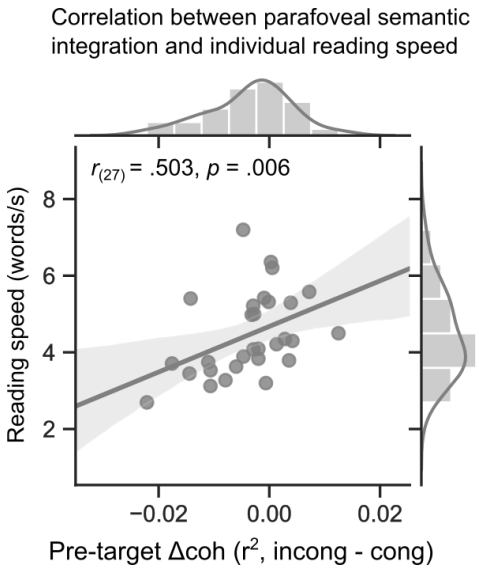

Fig. 4. Individual reading speed was related to the strength of parafoveal semantic integration. The reading speed was measured as the number of words read per second in the congruent sentences. The strength of parafoveal semantic integration was measured as the coherence difference during the pre-target fixations with respect to the incongruent and congruent target words. Each dot indicates one participant (n = 29, Spearman correlation). The shaded area represents the 95% CI.

Supplementary Materials for

Parafoveal semantic integration in natural reading

Yali Pan, Steven Frisson, Kara D. Federmeier, Ole Jensen*
*Correspondence: O.Jensen@bham.ac.uk

This PDF file includes:
Materials and Methods
Supplementary Text
Table S1
fig. S1
fig. S2
fig. S3
References (1-16)

Materials and Methods

Participants

We recruited 36 native English speakers (24 females, 22.5 ± 2.8 years old, mean ± SD) with normal or corrected-to-normal vision. All participants are right-handed and without any neurological history or language disorder diagnosis. Two participants were excluded from analysis due to poor eye tracking or falling asleep during the recordings, which left 34 participants (23 females). The study was approved by the University of Birmingham Ethics Committee. The informed consent form was signed by all participants after the nature and possible consequences of the studies were explained. Participants received £15 per hour or course credits as compensation for their participation.

Stimuli

In total participants read 277 sentences, of which 117 sentences were fillers from a published paper (*1*). The filler sentences were all plausible and were included to make sure the incongruent sentences were less than one third of the sentence set. We constructed the remaining 160 sentences with 80 pairs of target words. In all sentences the context was low constraint, i.e., none of the target words could be predicted by the prior context (see pre-tests below for details). The target word in each sentence was either incongruent or congruent with the sentence. By swapping the target words within a pair of sentences, we created four sentences: two congruent and two incongruent. These were then counterbalanced over participants. In this way, we counterbalanced across lexical characteristics of the target words and characteristics of the sentence frames within each pair. Each participants read one version of the sentence set (A or B). For example, for the target pair *brother/jacket*, one participant read them in the congruent condition in the sentence set version A; while another participant read them in the incongruent condition in version B (see below, targets are in italic type for illustration, but in normal type in the real experiment).

A. Last night, my lazy *brother* came to the party one minute before it was over.

Lily says this blue *jacket* will be a big fashion trend this fall.

B. Last night, my lazy *jacket* came to the party one minute before it was over.

Lily says this blue *brother* will be a big fashion trend this fall

For all sentences, the pre-target words were adjectives, and the target words were nouns (for detailed characteristics of the words please see Supplementary Table S1). The word length of pre-target words was from 4 to 8 letters, while for target words was from 4 to 7 letters. The sentences were no longer than 15 words or 85 letters. The target words were embedded somewhere in the middle of each sentence and were never the first three or the last three words in a sentence. Please see Supplementary Text for the full list of the sentence sets that were used in the current study.

Pre-tests for target predictability and sentence plausibility

We recruited native English speakers for two behavioural pre-tests of the sentence sets. These participants did not participate in the MEG session.

Predictability

We carried out a cloze test to estimate the predictability of the target words and the contextual constraint of the sentences. Participants read sentence fragments consisting of the experimental materials up to but not including the target words. Then participants were asked to write down the first word that came to mind that could continue the sentence (no need to complete the whole sentence). Example:

1. Last night, my lazy _____
2. Lily says this blue _____

The predictability of a word was estimated as the percentage of participants who wrote down exactly this word in the cloze test. A target word with less than 10% predictability was deemed to be not predicted by the sentence context. In addition, sentences for which no word was predicted with 50% or greater probability were low constraint. Twenty participants (6 males, 24.2 ± 2.0 years old, mean ± SD) took part in the first round of pre-test. Eight sentences were replaced with new sentences because the target words were too predictable and/or the sentence was too constraining. We then conducted a second round of the predictability test with 21 new participants (7 males, 25.0 ± 6.0 years old). None of the target words in this final set were predictable (2.3% ± 4.8%, mean ± SD), and all the sentence contexts were low constraint (25.2% ± 11.8%).

Plausibility

Two groups of participants were instructed to rate how plausible (or acceptable) each sentence was in the sentence set version A or B separately. Plausibility was rated on a 7-point scale with plausibility increasing from point 1 to 7. Sentences in the experiment were designed to be either highly implausible (the incongruent condition) or highly plausible (the congruent condition). To occupy the full range of the scale, we constructed 70 filler sentences with middle plausibility (e.g., sentence 1 below). In this example, sentences 2 and 3 were the incongruent and congruent sentences from the experiment.

	Implausible 1	2	3	4	5	6	Plausible 7
1. Kate said that she saw lots of stars twinkling in the sky at noon.	1	2	3	4	5	6	7
2. Lily says this blue brother will be a big fashion trend this fall.	1	2	3	4	5	6	7
3. Little Jimmy picked up a box and put some coins inside of it.	1	2	3	4	5	6	7
……	1	2	3	4	5	6	7

For version A we recruited 27 participants (4 males, 22.8 ± 6.1 years old, mean ± SD): The plausibility rating for the incongruent sentences was 2.08 ± 0.79 (mean ± SD); while for the congruent sentences was 6.18 ± 0.56. For sentence set version B we recruited 22 participants (4 males, 21.1 ± 2.3 years old, one invalid dataset due to incomplete responses): The plausibility rating was 1.81 ± 0.41 (mean ± SD) for the sentences in the incongruent

condition and 6.15 ± 0.47 for the sentences in the congruent condition. These results showed that in both versions of the sentences set, incongruent sentences were viewed as highly implausible and congruent sentences as highly plausible.

Experimental procedure

Participants were seated 145 cm away from the projection screen in a dimly lit magnetically shielded room. The MEG gantry was set at 60 degrees upright and covered the participant's whole head. We programmed in Psychophysics Toolbox -3(*2*) to present the one-line sentences on a middle-grey screen (RGB [128 128 128]). All words were displayed in black (RGB [0 0 0]) with an equal-spaced Courier New font. The font size was 20 and the font type was bold so that each letter and space occupied 0.316 visual degrees. The visual angle of the whole sentence was no longer than 27 visual degrees in the horizontal direction. The sentence set was divided into 5 blocks, each of which took about 7 minutes. There was a break of at least 1 minute between blocks and participants pressed a button to continue the experiment at any time afterwards. Participants were instructed to read each sentence silently at their own pace and to keep their heads and body as stable as possible during the MEG session. Eye movements were acquired during the whole session. In total, the experiment took no longer than 55 minutes.

Within a trial, first was a fixation cross presented at the centre of a middle-grey screen for 1.2 – 1.6 s. This was followed by a black square with a radius of 1 degree of visual angle. This square was placed at the vertical centre, 2 degrees of visual angle away from the left edge of the screen. Participants had to gaze at this black 'starting square' for at least 0.2 s to trigger the onset of the sentence presentation. Afterwards, the sentence would start from the location of the square (Fig. 1). The sentence was presented with an 'ending square' 5 degrees of visual angle below the screen centre. The 'ending square' was the same size as the 'starting square' but in grey colour (RGB [64 64 64]). A gaze at this 'ending square' for at least 0.1s would end the presentation of the sentence. Then this trial ended with a blank middle-grey screen that lasted for 0.5s. Randomly, 12% of the trials were followed by a statement about the content of the sentence that was just presented, and participants needed to answer "True or "False" by pressing a button. For example, the statement for sentence 2 was "Lily has a prediction about the fashion trend in this fall", and the correct answer was "True". The statement for sentence 3 was "Little Jimmy didn't have a box", and the correct answer was "False". All participants read the sentences carefully as shown by the high accuracy of answering (96.3% ± 4.7%, mean ± SD).

Rapid invisible frequency tagging (RIFT)

Projection of the sentence stimuli

We projected the sentences from the stimulus computer screen in the experimenter room to the projection screen inside of the MEG room using a PROPixx DLP LED projector (VPixx Technologies Inc., Canada). The refresh rate of the PROPixx projector was up to 1440 Hz, while the refresh rate of the stimulus screen was only 120 Hz (1920 × 1200 pixels resolution). We displayed the sentence repeatedly in four quadrants of the stimulus computer screen. In each quadrant, the words were coded in three colour channels as RGB. The projector then interpreted these 12 colour channels (3 channels × 4 quadrants) as 12 individual grayscale frames, which were projected onto the projection screen in rapid succession. Therefore, the projection screen refreshed at 12 times the rate of the stimulus computer screen.

Flickering of the target word

We added a square patch underneath the target word to frequency tag the target word. The side length of the square patch was the width of the target word plus the spaces on both sides (2 to 3° visual angle). We flickered the patch by changing its luminance from black to white at a 60 Hz sinusoid (Fig. 1). To reduce the visibility of the patch edges across saccades, we applied a Gaussian smoothed transparent mask on top of the square patch. The mask was created by a two-dimensional Gaussian function:

$$\text{mask} = \exp\left(-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}\right) \qquad \text{Equation 1}$$

where, x and y are the mesh grid coordinates for the flickering patch, and σ is the x and y spread of the mask with σ = 0.02 degrees.

On average, the patch was perceived as middle-grey, the same colour as the background screen, which made it invisible to participants. The target word was still black, the same colour as the other words on the screen. To record the tagging signal, we attached a custom-made photodiode (Aalto NeuroImaging Centre, Finland) to the disk at the bottom right corner of the screen. The luminance of the disk varied the same as that of the flickering patch underneath the target word. The photodiode was plugged into the MEG system as an external channel.

Data acquisition

MEG

Brain data were acquired with a 306-sensor TRIUX Elekta Neuromag system, which consisted of 204 orthogonal planar gradiometers and 102 magnetometers (Elekta, Finland). After participants signed the consent form, we attached four head-position indicator coils (HPI coils) to their heads: two on the left and right mastoid bone, and two on the forehead with at least 3 cm distance in between. Afterwards, we used a Polhemus Fastrack electromagnetic digitizer system (Polhemus Inc, USA) to digitize the locations for three bony fiducial points: the nasion, left and right preauricular points. Then we digitized the four HPI coils. Furthermore, at least 200 extra points were acquired, which were distributed evenly and covered the whole scalp. These points were used later in the source analysis when spatially co-register the MEG hear model with individual structural MRI images. The sampling rate of the MEG system was 1,000 Hz. Data were band-pass filtered prior to sampling from 0.1 to 330 Hz to reduce aliasing effects.

Eye movements

We used an EyeLink 1000 Plus eye-tracker (long-range mount, SR Research Ltd, Canada) to track eye movements throughout the whole MEG session. The eye tracker was placed on a wooden table in front of the projection screen. The centre of the eye tracker was at the middle line of the projection screen, and the top of the eye tracker reached the bottom edge of the screen. The distance between the eye-tracker camera and the centre of the participant's eyes was 90 cm. We recorded the horizontal and vertical positions as well as the pupil size from the left eye, at a sampling rate of 1,000 Hz. Each session began with a nine-point calibration and validation test. The test was accepted if the eye-tracking error was below 1 visual degree both horizontally and vertically. During the session, we performed a one-point drift checking test every three trials and after the break between blocks. If the drift checking failed or the sentence presentation was unable to be triggered through gazing, a nine-point calibration and validation test was conducted again.

MRI

After MEG data acquisition, participants were asked to come to the laboratory another day to have an MRI image acquired. We acquired the T1-weighted structural MRI image using a 3-Tesla Siemens PRISMA scanner (TR = 2000 ms, TE = 2.01 ms, TI = 880 ms, flip angle = 8 degrees, FOV = 256×256×208 mm, 1 mm isotropic voxel). For 11 participants who dropped out of the MRI acquisition, the MNI template brain (Montreal, Quebec, Canada) was used instead in the source analysis later.

Eye movement data analysis

We extracted the fixation onset events from the EyeLink output file. The EyeLink parsed fixation events based on the online detection of saccade onset using the following parameters: the motion threshold as 0.1 degrees, the velocity threshold as 30 degrees/sec, and the acceleration threshold as 8000 degrees/sec2. These conservative settings were suggested by the EyeLink user manual for reading studies, which can prevent false saccade reports and reduce the number of micro-saccades and lengthen fixation durations.

Then, only the fixation that first landed on a given word was selected. The first fixation durations were averaged within the incongruent and congruent conditions for pre-target and target words. Pairwise, two-sided *t*-test were conducted to the first fixation durations of pre-target and target words separately. In addition to the early eye movement measure of the first fixation duration, we also conducted *t*-tests for two later eye movement measures. The likelihood of refixation was measured as the proportion of trials on which there was at least one saccade that regressed back to that word. The total gaze duration was the sum of all fixations on a given word, including those fixations during regression of re-reading.

MEG data analyses

The data analyses were performed in MATLAB R2020a (Mathworks Inc, USA) by using the FieldTrip(*3*) toolbox (version 20200220) and custom-made scripts.

Pre-processing

We first band-pass filtered the MEG data from 0.5 to 100 Hz using phase preserving two-pass Butterworth filters. Then, we detrended the data to factor out the slow drifts and the linear trends. Malfunctioning sensors were removed based on inspecting the data quality during online recording (0 to 2 sensors per participant). Afterwards, the data were decomposed into independent components using an independent component analysis (ICA)(*4*). The number of components was the same as the number of good MEG sensors in the dataset (306 or less). We only removed bad components that related to eye blinks, eye movements, and heartbeat by visually inspecting the components (3.4 ± 0.7 components per participant, M ± SD, range from 2 to 5 components).

MEG segments were extracted from -0.5 to 0.5 s intervals aligned with the first fixation onset of the pre-target and target words respectively. For how to define the fixation onset to a word, please see Eye movement data analysis above. Segments with fixation durations shorter than 0.08s or longer than 1 s were discarded. We also extracted 1 s long baseline segments, which were aligned with the cross-fixation onset before the sentence presentation. We manually inspected all segments to further identify and remove segments that were contaminated by muscle or movement artefacts.

Coherence calculation

We calculated the coherence between the MEG sensors and the photodiode (i.e., the tagging signal) to quantify the tagging responses. The amplitude of photodiode channel was normalized across each segment. To estimate the coherence spectrum in the frequency domain over time, we filtered the segments using hamming tapered Butterworth bandpass filters (4th order, phase preserving, two-pass). The frequency of interest was from 40 to 80 Hz in a step of 2 Hz. For each centre frequency point, the spectral smoothing was ± 5 Hz. For example, the filter frequency range for 60 Hz was from 55 to 65 Hz. We performed a Hilbert transform to obtain the analytic signals for each centre frequency point, which then was used to estimate the coherence:

$$\mathrm{coh}(t) = \frac{\left| n^{-1} \sum_{j=1}^{n} m_{x_j}(t) m_{y_j}(t) e^{i\emptyset_{xy_j}(t)} \right|^2}{n^{-1} \sum_{j=1}^{n} m_{x_j}(t)^2 \, m_{y_j}(t)^2} \qquad \text{Equation 2}$$

where n is the number of trials. For the time point t in the trial j, $m_x(t)$ and $m_y(t)$ are the time-varying magnitude of the analytic signals from a MEG sensor (x) and the photodiode (y) respectively, $\emptyset_{xy}(t)$ is the phase difference as a function of time.

Selection for the RIFT response sensors

MEG sensors that showed significantly stronger coherence at the 60 Hz during the pre-target segments than the baseline segments were selected as the RIFT response sensors. We used a non-parametric Monte-Carlo method (*5*) to estimate the statistical significance. The pre-target segments were constructed by pooling the target congruency conditions together. Several previous RIFT studies from our lab observed robust tagging responses from the visual cortex for flicker above 50 Hz (*6–9*). Thus, this sensor selection procedure was confined to the MEG sensors in the visual cortex (52 planar sensors). Here, the pre-target segments and baseline segments were treated as two conditions. For each combination of MEG sensor and photodiode channel, coherence at 60 Hz was estimated over trials for the pre-target and baseline conditions separately. Then, we calculated the z-statistic value for the coherence difference between pre-target and baseline using the following equation (for details please see Maris et al., 2007(*5*)):

$$\mathrm{Z} = \frac{(\tanh^{-1}(|\mathrm{coh}_1|) - \mathrm{bias}_1) - (\tanh^{-1}(|\mathrm{coh}_2|) - \mathrm{bias}_2)}{\sqrt{\mathrm{bias}_1 + \mathrm{bias}_2}}, \qquad \text{Equation 3}$$

$$\mathrm{bias}_1 = \frac{1}{2n_1 - 2}, \ \mathrm{bias}_2 = \frac{1}{2n_2 - 2}$$

where coh_1 and coh_2 denote the coherence value for pre-target and baseline segments, bias_1 and bias_2 is the term used to correct for the bias from trial numbers of pre-target (n_1) and baseline condition (n_2). All trials from the pre-target and baseline conditions were used.

After obtaining the z statistic value for the empirical coherence difference, we ran a permutation procedure to estimate the statistical significance of this comparison. We randomly shuffled the trial labels between pre-target and baseline conditions 5,000 times. During each permutation, coherence was computed for both conditions (with shuffled labels), then entered Equation 3 to obtain a z score for the coherence difference. After all randomizations were performed, the resulting z-values established the null distribution. Since a tagging response sensor was supposed to have stronger coherence during the pre-target segments compared with the baseline segments, the statistical test was right-sided. If the z-value of the empirical coherence difference was larger than 99% of z-values in the null

distribution, this sensor was selected as the RIFT response sensor (right-sided, $p = 0.01$). For each participant, the coherence values were averaged over all sensors with significant tagging response to obtain an averaged coherence for further analyses.

Coherence response curve

We first extracted MEG segments for the words N-4, N-3, N-2, N+1, N+2, and N+3 following the same procedure described in the pre-processing when extracted MEG segments for the pre-target (N-1) and target words (N). All segments were 1 s long, aligned with the first fixation onset to the word. Then, we calculated the coherence at 60 Hz during these segments for participants who have RIFT response sensors (n = 29). Next for each participant, the 60 Hz coherence was first averaged over the RIFT response sensors, then averaged within a time window of [0 0.2] s (the averaged fixation duration for words). We got an averaged 60 Hz coherence for the word at each position. We also got the 60 Hz coherence for the baseline interval averaged over [0 0.2] s, aligned with the cross-fixation onset. Then a pairwise *t* test was performed between the baseline coherence and the coherence at each word position.

Coherence comparison between conditions

The coherence comparison analyses were only conducted for the participants who had sensors with a reliable tagging response (n = 29). To avoid any bias from trial numbers, an equal number of trials under the different congruency conditions was entered the coherence analysis per participant. We randomly discarded the redundant trials from the condition that had more trials for both the pre-target and target segments.

To compare the pre-target coherence amplitude between conditions, the coherence values at 60 Hz were averaged across the minimum fixation duration of all pre-target words. The time window for averaging was defined for each participant so that the coherence signal from the target fixation was not involved. Similarly, we averaged the 60 Hz coherence for the target segments over the minimum target fixation duration. Then, a two-sided pairwise Student's *t*-test was performed to estimate the statistical significance of the coherence difference as shown in Fig. 3B and fig. S1B.

To assess the coherence onset latency difference between conditions, we used a leave-one-out Jackknife-based method (*10*). We extracted the 60 Hz coherence during the 1 s long pre-target segments for each participant. Then, during each iteration of participants, we randomly chose and left out one participant. For the remaining participants, coherences were calculated for the incongruent and congruent target conditions. Then, the coherence was averaged over the remaining participants to estimate the onset latency for both conditions. Here, the onset latency was defined as the time point when the averaged coherence value reached its half-maximum ($\mathrm{coh}_{\mathrm{min}} + (\mathrm{coh}_{\mathrm{max}} - \mathrm{coh}_{\mathrm{min}})/2$). We computed the onset latency difference by subtracting the onset latency for the incongruent target condition from the congruent condition. After all iterations, onset latency differences from all these subsamples were pooled together to estimate a standard error (S_D) using the following equation:

$$S_D = \sqrt{\frac{n-1}{n} \cdot \sum_{i=1}^{n} \left(D_{-i} - \bar{J}\right)^2} \qquad \text{Equation 4}$$

where the $\bar{J}$ is the averaged onset latency difference over all the subsamples, D_{-i} is the coherence difference obtained from the subsample when participant i was left out, n is the number of participants. We also computed the onset latency difference from the overall sample set (without leaving any participant out) and divided it by the S_D to obtain its *t*-value. A standard t table (pairwise, two-tailed) provided the statistical significance for the coherence onset latency difference between the incongruent and congruent target conditions. This procedure was conducted for both the pre-target and target segments as shown in Fig. 4D and 4E.

Source analysis for RIFT

We used a beamforming-based approach, Dynamic Imaging Coherent Sources (DICS) (*11*), to estimate the neural sources that generated the responses to RIFT. The DICS technique was applied to the pre-target segments regardless of the target congruency conditions, with a focus of 60 Hz in the frequency domain. In this source analysis, only participants with robust tagging responses were included (n = 29).

First, we constructed the semi-realistic head model, where spherical harmonic functions were used to fit the brain surface (*12*). We aligned the individual structural MRI image with the head shape that was digitized during the MEG session. This was done by spatially co-registering the three fiducial anatomical markers (nasion, left and right ear canal) and the extra points that covered the whole scalp. For participants whose MRI image was unavailable, the MNI template brain was used instead. The aligned MRI image was segmented into a grid, which was used to prepare the single-shell head model.

Then, we constructed the individual source model by inverse-warping a 5 mm spaced regular grid in the MNI template space to each participant's segmented MRI image. We got the regular grid from the Fieldtrip template folder, which was constructed before doing the source analysis. In this way, the beamformer spatial filters were constructed on the regular grid that mapped to the MNI template space. Even though after this warping procedure, grid points in the individual native space were not evenly spaced anymore, the homologues grid points across participants were located at the same location in the normalized template space. Thus, the reconstructed sources can be directly averaged across participants on the group level.

Next, the Cross-Spectral Density (CSD) matrix was calculated at 60 Hz for both the pre-target and baseline segments. The CSD matrix was constructed for all possible combinations between the MEG sensors and the photodiode channel. No regularisation was performed to the CSD matrices (lambda = 0).

Finally, a common spatial filter was computed based on the individual single-shell head model, source model, and CSD matrices. This spatial filter was applied to both the pre-target and baseline CSD matrices for calculating the 60 Hz coherence. This was done by normalizing the magnitude of the summed CSD between the MEG sensor and the photodiode channel by their respective power. After the grand average over participants, the relative change for pre-target coherence was estimated as the following formula, $(coh_{pretarget} - coh_{baseline})/coh_{baseline}$.

Fixation-related fields (FRF) analysis

FRF analysis was conducted for all participants (n = 34) during the pre-target and target segments. All segments were 1 s long, with 0.5 s before and after the fixation onset to the

word (-0.5 to 0.5 s). For each participant, the same number of trials were randomly selected based on the minimum trial number across the target congruency conditions. This was done for pre-target and target segments separately to avoid any bias from the trial number to the FRF analysis estimated for the combined planar gradient.

First, we low-pass filtered the data under 35 Hz using phase preserving, two-pass Butterworth filters (4th order with a hamming taper). We did a baseline correction by subtracting the data with a time window of -0.2 to 0 s. Then FRFs were estimated by averaging the segments. For each pair of planar gradient sensors at the same location, the combined FRFs were obtained as the root-mean-square of the FRFs from the horizontal and vertical sensors.

A cluster-based permutation test was conducted to estimate the target congruency effect for the pre-target and target segments. The time window of the permutation test was 0 to 0.5 s aligned with the fixation onset to the pre-target or target words. In each permutation loop, the labels for incongruent and congruent target conditions were randomly assigned to participants. Then for each sensor, a two-tailed pair-wise t-test was conducted to estimate the conditional difference. At least two significant neighbouring sensors (p = 0.05) were required to form a cluster. The neighbourhood layout for sensors was pre-defined by FieldTrip for the MEGIN system. We summed up the t-values within each cluster of sensors and then selected the maximum summed t-values to construct the null distribution. After 1000 permutations, we sorted all the summed t-values from the null distribution in ascending order. T-values at the 25th and the 975th position were chosen as the critical *t*-values for the negative ($\mathrm{FRFs}_{\mathrm{incongruent}} < \mathrm{FRFs}_{\mathrm{congruent}}$) and positive ($\mathrm{FRFs}_{\mathrm{incongruent}} > \mathrm{FRFs}_{\mathrm{congruent}}$) clusters. Then we compared the summed t-values from the experiment data with these critical t-values from the shuffled data to get the statistical significance of the sensor clusters.

Statistical information

All the *t* tests in this study were pairwise student's *t*-tests and conducted in R (*13*).

Supplementary Text

Experimental sentence set

Here we share all 160 sentences embedded with congruent target words. For sentence set version A, we swapped the target words within each pair for sentences 1 to 80 and made them incongruent, while sentences 81 to 160 were kept congruent. For sentence set version B, target words in sentences 1 to 80 were kept the same but target words in sentences 81 to 160 were swapped within each pair to make them incongruent. Then the sequence of the sentences was shuffled and make sure no more than 3 sentences in a row were in the same condition. For illustration, the target words are shown in italic type here, but they were in normal type in the experiment. For the 117 filler sentences, please see the Appendix in (*1*).

1. Last night, my lazy *brother* came to the party one minute before it was over.
2. Lily says this blue *jacket* will be a big fashion trend this fall.

3. This area has been populated by many *hikers* over the last year.
4. Little Jimmy picked up a box and put some *coins* inside of it.

5. Joey became an avid *student* during his adolescence.

6. He could only afford a cheap *ring* without a diamond for his fiancée.

7. This morning the noisy *kids* played happily in the backyard.
8. My parents had no firm *ideas* about what I should become.

9. The unfortunate pupil lost his beloved *pony* just before his birthday.
10. Experts say that the severe *crisis* will cause oil prices to triple.

11. The construction of this ancient *castle* cost a lot of money.
12. After the meeting, the anxious *worker* sighed in the hallway.

13. Peter's love for this sporting *match* inspired all his friends.
14. We could see from her torn *collar* that she had been in a fight.

15. With the help of his clever *friend* Jack, he made the first pot of gold.
16. I always like to order a filling *burger* from the local pub.

17. She looked at the tired *fireman* with a satisfied smile.
18. He felt relieved after completing the complex *scan* within an hour.

19. Scientists found a steep *boulder* sitting in the middle of the canyon.
20. Last week his friendly *cousin* passed out for no apparent reason.

21. They asked the selfish *maid* where her huge sums of money came from.
22. It took Tom a month to mend the broken *roof* all by himself.

23. Under stress, the crafty *boss* promised customers a full refund.
24. The cowboys hung the stolen *rifles* high up on the wall.

25. She submitted the crucial *file* that can prove her innocence.
26. According to history books, the proud *queen* never accepted any criticism.

27. Ana complained that the tall *herbs* behind the house had dried up.
28. He failed in his chosen *sport* with hopes of success fading with each effort.

29. She said that the corrupt *company* offered high salaries to young graduates.
30. In the last few years, the thick *bush* died back dramatically.

31. It turned out that the last- minute *trip* lasted for six hours.
32. Linda found that the slender *tree* dead from a pest infestation.

33. Suddenly, the warm *coffee* stained his brand new shirt.
34. Plenty of rain will make the vivid *flower* blossom well.

35. Jack became a humble *chef* specializing in French cuisine.
36. The young man's shiny *vehicle* vanished slowly out of sight.

37. To the north, the steep *hills* stretched for many miles.
38. Before sleeping, the nervous *colonel* smoked a cigarette.

39. Decades ago, that algae- covered *pond* was enough to irrigate the crops.
40. He saw the smart *player* throw the ball, causing chaos among the opposition.

41. In recent days, the cruel *murder* has scared citizens from going out.
42. Mary told me that the light *cream* was low in fat but hard to whip.

43. News said that the painful *disease* would continue to affect many children.
44. The boy found his lost *ball* under the tree and stopped crying at once.

45. Politicians hated the brief *report* criticizing the government's incompetence.
46. My favourite gift is the shiny *drone* from my dad last year.

47. Every year, the sandy *shore* attracts thousands of tourists.
48. After taking a deep breath, the junior *officer* entered the room.

49. Last week, the caring *family* rescued a stray dog and kept it as a pet.
50. During the air raid, the spacious public *plaza* happened to be ruined.

51. With his sharp criticism, the young *actor* annoyed his agent as usual.
52. Laura was told that the sudden *storm* delayed the bus for two days.

53. Lily said that the vacant *cottage* belonged to her grandparents.
54. In the small house, a comic *picture* adorned the reception room.

55. Facing the lion, the brave *hunter* showed no fear.
56. Out of repair, the rattling *engine* was about to be scrapped.

57. Jack had to admit that this planned *visit* turned out to be embarrassing.
58. Tom admired the way his devoted *aunt* always volunteers on weekends.

59. Rob felt that the brief *letter* from his wife expressed a hint of sadness.
60. Alone at home, the tired *clerk* cooked a beef patty.

61. After the surgery, Rob's poor *health* left him barely able to get out of bed.
62. Sam's train arrived before *dusk* and we were able to give him a ride home.

63. She gave the dog a quick *bath* after they came back from the outside.
64. Michael made a mean *joke* about Boris Johnson's hair.

65. They didn't realize the harsh *impact* that their products could have.
66. In the museum, we saw the golden *crown* that belonged to the first king.

67. Bill is a superb *partner* because he is easy to get along with.
68. The explorer made his way through the gloomy *night* with a small torch.

69. The TV show was an obvious *flop* after the actress joined the cast.
70. On rainy days, the careful *lady* reminded herself to go slowly.

71. Jane complained that her white *kitten* hadn't come home for two days.
72. I guess no one can solve the hard *problem* without outside help.

73. The new event was such a huge *failure* that people kept talking about it.
74. Before committing the crime, the anxious *suspect* drank a lot of alcohol.

75. Toby kept his money in a small *shed* because he lived on a farm.
76. They noticed the young *deer* eating acorns in the forest.

77. Amy wanted some more of the sliced *pear* for afternoon snack.
78. Laura went down to the narrow *canal* to watch the boats.

79. I wondered if the noisy *club* would be a good place for the bachelorette party.
80. Tara always has an acute *pain* in her tooth after eating ice cream.

81. Before the war, the brave *general* assembled an army.
82. She began to slice up a large *potato* for the dinner.

83. The child had a large *face* with big, expressive eyes.
84. Last night, there was a dense *mist* when they left the cinema.

85. Marla enjoyed seeing the chubby *cats* playing with each other.
86. Jim entered the giant *court* to try out for the basketball team.

87. They visited the antique *chapel* before booking their wedding.
88. After the defeat, the crazy *fans* kept cursing and crying.

89. She approached the rusty *gate* before realizing it was locked.
90. Many people like to eat crispy *toast* with their morning coffee at breakfast.

91. They stepped into the messy *garage* that had high wooden shelves.
92. We watched the large hungry *hawk* swoop down to get the poor chicken.

93. Alexandra used a short *hammer* when she created the stone statue.
94. Ruth visited the public *museum* that she had read about all these years.

95. We'd better buy some tasty *chips* before we watch the big game.
96. Historians believe the rousing *speech* heralded the start of the revolution.

97. Under the tree, there is a little *hare* running happily.
98. In the darkness, only the misty *moon* lit up the street.

99. The prince inherited the supreme *power* from the late king.
100. Look over there, a fluffy *sheep* seems to be lost.

101. The man was a young *teacher* who always worked late into the night.
102. As for this scandal, Jo has a clear *opinion* but she won't say it.

103. They had no idea that the blue *liquid* shrinks all woollen clothes.
104. The report was sent to the honest *justice* three days before the trial.

105. Every night, this tired *captain* drank wine before going to sleep.

106. The shopkeeper said the metal *bottle* would sell well this year.

107. For the locals, the salt *lake* triggered a political issue.
108. Near the small brook, a hungry *animal* hunts quietly for hours.

109. Every night, this deep *secret* makes the pianist toss and turn.
110. In the lab, a young *surgeon* examined the victim's body.

111. Sadly, the lonely *poet* died before he could finish his last poem.
112. Due to the moist weather, the wheat *flour* became mouldy quickly.

113. Sue's colleagues say that her warm *heart* makes everyone like her.
114. On the wall, the small green *screen* shows the room temperature precisely.

115. To his surprise, the yummy *dish* was not expensive.
116. Eventually, the greedy *nanny* disclosed all the details about this affair.

117. The sight of the cotton *factory* was something to behold.
118. It was obvious that the weak *patient* was getting weaker day by day.

119. In the past month alone, the gentle *scholar* published five papers.
120. The filthy and rusty *pots* made the food taste terrible.

121. Just after dawn, an armed *ship* approached the pretty lagoon slowly.
122. At last, she found the wool *skirt* hanging in the wardrobe.

123. Villagers said that the newly built *school* was well equipped.
124. In the downtown market, the agitated *crowd* began the parade.

125. Suzy really likes eating *sugar* because she wasn't allowed to eat it as a kid.
126. Ali said he really enjoyed modern *music* when he was at college.

127. Tina wants a spacious *yard* because she likes to lie on the grass and read.
128. In an open field, the violent *chief* executed prisoners with a gun.

129. After working overtime for a month, the wronged *manager* wanted to jump ship.
130. In the Stone Age, the spiny *grass* prevailed over the land.

131. The sparrow was being chased by some fluffy *hens* under the hot sun.
132. When the ball was scored, they tapped their *cups* to show their joy.

133. They recorded the details of the stolen *cars* carefully on a spreadsheet.
134. The poor boy stood in the snow with bruised *legs* and cried sadly.

135. Little Roy likes to play with the plastic *bricks* at the Lego store.
136. It was said that the honest *lawyer* convened the committee meeting.

137. I learned about the muddy *trail* through a friend on the last hike.
138. Just now, the calm *jury* delivered a guilty verdict in this notorious case.

139. The holiday was neglected by this busy *parent* but her son was used to it.
140. This widely spread *story* reflected the distortion of human nature.

141. We are meeting at the newly built *airport* tonight for our trip to Europe.
142. Every day before leaving work, the tall *editor* cleans her desk.

143. Nobody knew when the excited *puppy* urinated on the floor.
144. Everyone knows that entire *area* has restricted access.

145. The man's cunning *excuse* relieved him of the fine.
146. Roy repaired the broken *truck* over the weekend.

147. Ana was glad that the gentle *nurse* said her little boy was out of danger.
148. In the company, the annual *meeting* marks the end of a year's hard work.

149. He carefully placed the sharp *sword* down after the fight.
150. She found an empty *desk* where she could put her computer.

151. They danced a slow *tango* together after dinner.
152. Steph noticed a torn *note* and looked for the other half.

153. Mindy's dog has a strange *smell* and likes to bark a lot.
154. Patty likes to cut some pink *tape* to decorate her notebooks.

155. We could hear the angry *priest* shouting at the little girl.
156. David was happy to receive a nice *card* from his daughter at Christmas.

157. She always meets the same happy *couple* when she walks in the park.
158. The stray dog lives in a hidden *hole* that protects it from the cold weather.

159. Jack failed to submit his concise *paper* before the deadline.
160. I have heard that the young *baker* makes the best baguettes in town.

Figures and tables in the supplementary material

Table S1 Characteristics of pre-target, target, and post-target words

	Pre-target	Target	Post-target
Word frequency	124.5 (310.9)	62.2 (77.0)	3619.8 (6725.2)
Word Length	5.5 (1.1)	5.3 (1.1)	5.4 (2.0)
Position in the sentence	5.4 (1.3)	6.4 (1.3)	7.4 (1.3)

Note. Word frequency is reported as the total CELEX frequency per million (*14*). Word length is the number of letters in a given word. Position in the sentence refers to the location in the sequence of words where a given word is presented. The number of words in each sentence is 11.6 ± 1.7 (mean ±

SD). All values shown here are mean with standard deviations in the parentheses. Note that the pre-target, target, and post-target words that are reported here are regardless of the conditions since they are the same for both conditions over participants.

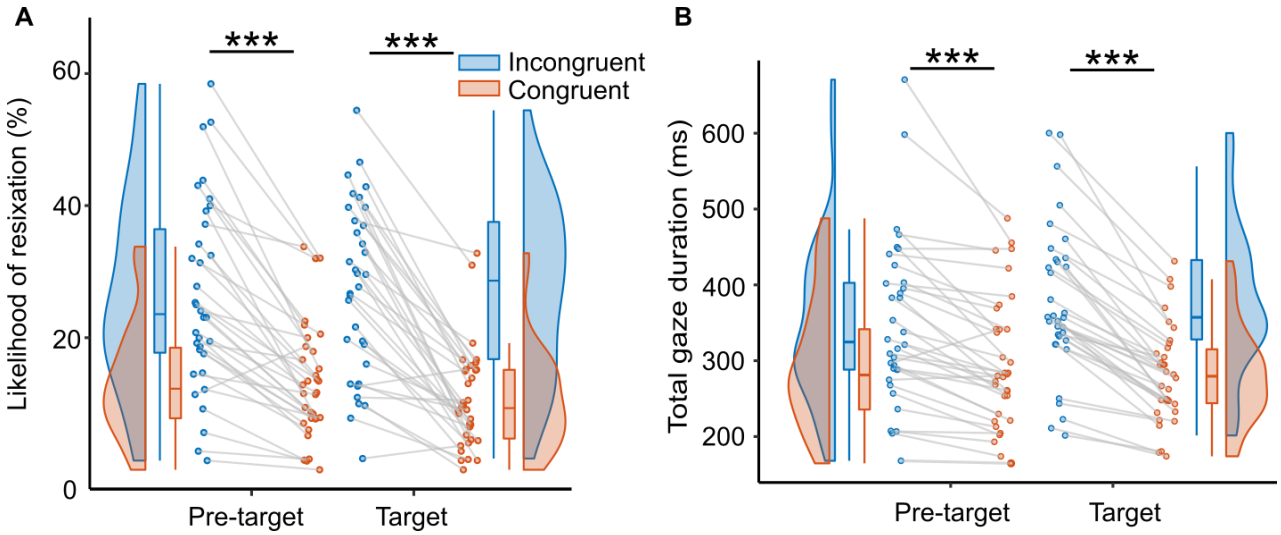

fig. S1. The likelihood of refixation and total gaze duration of eye movement data. (**A**) The likelihood of refixation into a word was defined as the proportion of trials that have at least one regression from a later part of the sentence back to that word. We found that when the target words were contextually incongruent with the sentence compared with congruent, there were significantly higher probability of regression into pre-target words ($t_{(33)} = 7.83$, $p = 5.04 \times 10^{-9}$, $d = 1.34$, two-sided pairwise *t*-test) and target words ($t_{(33)} = 9.13$, $p = 1.49 \times 10^{-10}$, $d = 1.57$, two-sided pairwise *t*-test). Each dot indicates one participant. (**B**) The total gaze duration was defined as the sum of all fixations on a given word, including those fixations during regression of re-reading. Significantly longer total gaze durations were found for pre-target words ($t_{(33)} = 5.78$, $p = 1.86 \times 10^{-6}$, $d = 0.99$, two-sided pairwise *t*-test) and target words ($t_{(33)} = 10.55$, $p = 4.20 \times 10^{-12}$, $d = 1.81$, two-sided pairwise *t*-test) when the target words were incongruent with the context compared with congruent. ***$p < .001$; n.s., not statistically significant.

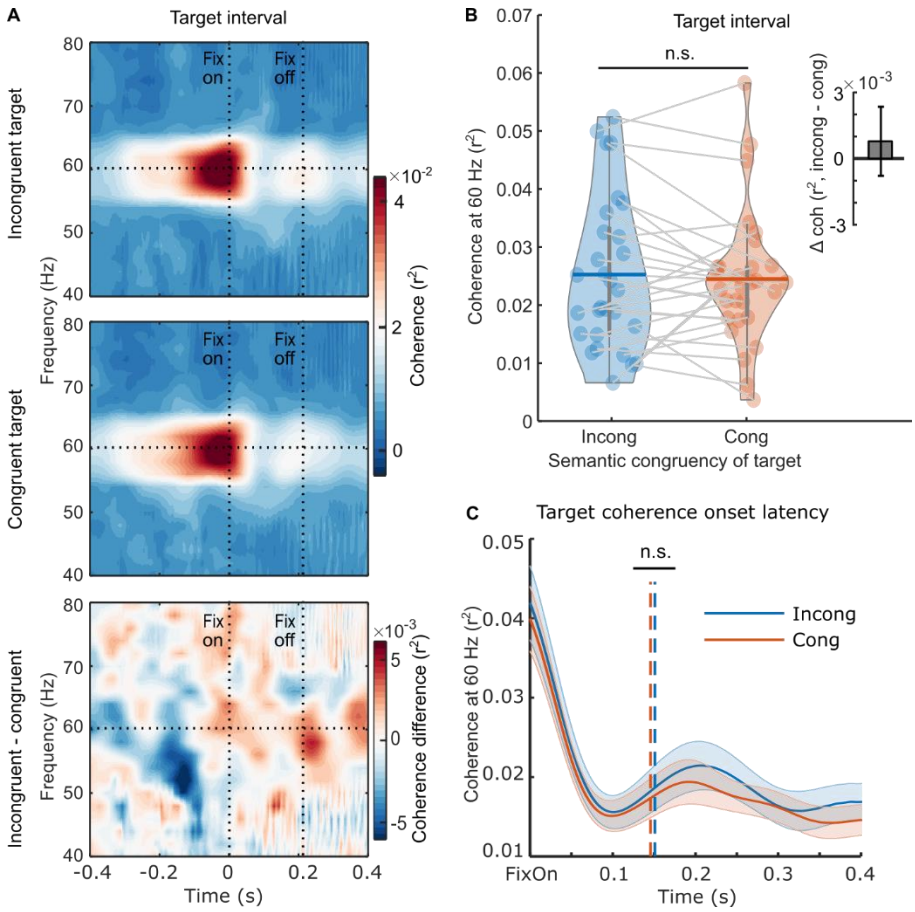

fig. S2. Neural responses to the rapid invisible frequency tagging (RIFT) during the target intervals. (**A**) The target coherence spectrum averaged over the RIFT response sensors at the group level (n = 29) when the target words were incongruent with the sentence context (top panel), or congruent with the sentence context (middle panel); the bottom panel shows the difference between the two conditions. The horizontal line indicated the tagging frequency at 60 Hz. The two vertical lines indicated the first fixation onset of the target words and the averaged fixation offset. (**B**) We averaged the 60 Hz coherence within the minimum target fixation duration over participants (97.6 ± 15.7 ms, M ± SD). No significant modulation of the target congruency was found for the magnitude of coherence ($t_{(28)}$ = .498, p = .622, pairwise, two-sided). Each dot indicates one participant, the horizontal lines inside of the violins indicate the mean values. The upright inserted figure showed the target coherence difference over participants with the error bar as SE. (**C**) A jackknife-based method was used to calculate the onset latency of the averaged coherence at the group level. The onset latency refers to the time when the coherence curve reached its half maximum, denoted by the dotted lines. The onset latency of 60 Hz coherence during the target intervals was not modulated by the congruency of the target words ($t_{(28)}$ = -.280, p = .782, pairwise, two-sided).

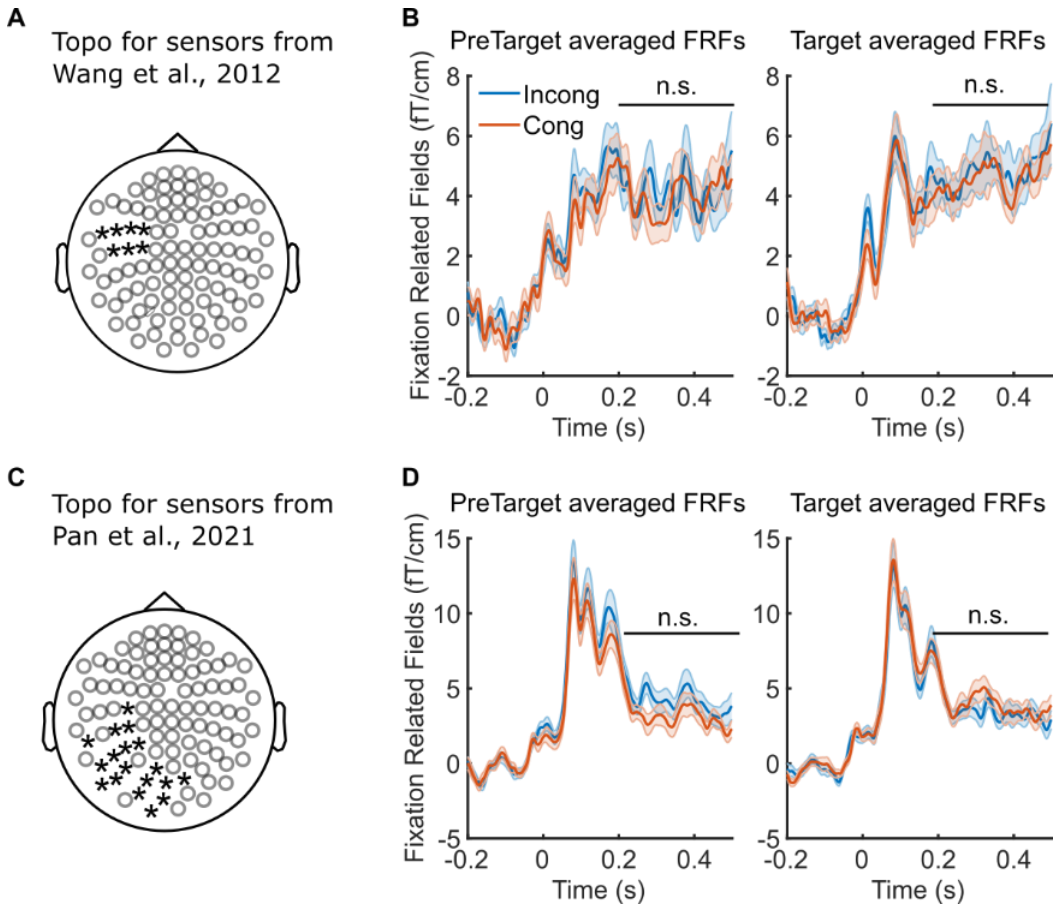

fig. S3. No semantic violation effect was observed with fixation-related fields (FRFs) in the natural reading task. FRFs were calculated during the pre-target and target intervals, and no cluster of sensors was found to be significant between the congruent and incongruent conditions of target words (n = 34, cluster-based permutation test with a significance level of 0.05). To show this null finding, we averaged the FRFs over sensors from two published MEG studies. (**A**) The topography of significant sensors (denoted by *) from Wang et al., 2012 (*15*), where a semantic congruency mN400 effect was found in a passive reading task. The averaged FRFs curves over the selected sensors are shown in (**B**) during the pre-target (left panel) and the target intervals (right panel). (**C**) The topography of significant sensors that showed a word frequency related mN400 effect in a natural reading task (Pan et al., 2021 (*16*)). (**D**) The averaged FRFs over the selected sensors in (**C**) during the pre-target (left panel) and the target intervals (right panel). n.s., not statistically significant.

Supplementary references

1. S. J. White, Eye movement control during reading: Effects of word frequency and orthographic familiarity. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 205–223 (2008).

2. M. Kleiner, D. Brainard, D. Pelli, What's new in Psychtoolbox-3? **14** (2007).

3. R. Oostenveld, P. Fries, E. Maris, J.-M. Schoffelen, FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011** (2011), doi:10.1155/2011/156869.

4. S. Ikeda, K. Toyama, Independent component analysis for noisy data - MEG data analysis. *Neural Networks*. **13**, 1063–1074 (2000).

5. E. Maris, J. M. Schoffelen, P. Fries, Nonparametric statistical testing of coherence differences. *J. Neurosci. Methods*. **163**, 161–175 (2007).

6. A. Zhigalov, J. D. Herring, J. Herpers, T. O. Bergmann, O. Jensen, Probing cortical excitability using rapid frequency tagging. *Neuroimage*. **195**, 59–66 (2019).

7. A. Zhigalov, O. Jensen, Alpha oscillations do not implement gain control in early visual cortex but rather gating in parieto-occipital regions. *Hum. Brain Mapp.* **41**, 5176–5186 (2020).

8. L. Drijvers, O. Jensen, E. Spaak, Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Hum. Brain Mapp.* **42**, 1138–1152 (2021).

9. K. Duecker, T. P. Gutteling, C. S. Herrmann, O. Jensen, No Evidence for Entrainment: Endogenous Gamma Oscillations and Rhythmic Flicker Responses Coexist in Visual Cortex. *J. Neurosci.* **41**, 6684–6698 (2021).

10. J. Miller, T. Patterson, R. Ulrich, Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*. **35**, 99–115 (1998).

11. J. Gross, J. Kujala, M. Hämäläinen, L. Timmermann, A. Schnitzler, R. Salmelin, Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 694–699 (2001).

12. G. Nolte, The magnetic lead field theorem in the quasi-static approximation and its use for magnetoenchephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* **48**, 3637–3652 (2003).

13. R. C. Team, R: A language and environment for statistical computing. **201** (2013).

14. C. J. Davis, N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behav. Res. Methods*. **37**, 65–70 (2005).

15. L. Wang, O. Jensen, D. van den Brink, N. Weder, J.-M. Schoffelen, L. Magyari, P. Hagoort, M. Bastiaansen, Beta oscillations relate to the N400m during language comprehension. *Hum. Brain Mapp.* **33**, 2898–2912 (2012).

16. Y. Pan, S. Frisson, O. Jensen, Neural evidence for lexical parafoveal processing. *Nat. Commun. 2021 121*. **12**, 1–9 (2021).