Working Memory Reactivation Across Embedded Structures during Natural Language Comprehension

Jiaqi Li (jiaqi.li@psych.ox.ac.uk)
Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging,
Department of Psychiatry, University of Oxford, Oxford, United Kingdom
Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

Yali Pan (y.pan.1@bham.ac.uk)
Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham,
United Kingdom
Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging,
Department of Psychiatry, University of Oxford, Oxford, United Kingdom

Hyojin Park (h.park@bham.ac.uk)
Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham,
United Kingdom

Huan Luo (huan.luo@pku.edu.cn)
School of Psychological and Cognitive Sciences, Peking University, Beijing, China
PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China
Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

Ole Jensen (ole.jensen@psych.ox.ac.uk)
Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging,
Department of Psychiatry, University of Oxford, Oxford, United Kingdom
Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

Abstract

A key feature of human language is recursion, involving hierarchical embedding of clauses– a process dependent on working memory (WM). During speech comprehension, listeners must maintain earlier linguistic elements for later integration. Even processing adjacent words into phrases requires WM resources (Desbordes et al., 2024), whereas the integration across embedded structures depends more on the flexibility of the WM system. However, the neural mechanisms underlying how WM supports the processing of complex recursive linguistic structures remain unclear. We constructed English sentences with embedded language structures and recorded magnetoencephalography (MEG) signals while 34 native speakers listened to these sentences. Neural decoding results demonstrate that during embedded structure processing, previously encoded information is stored in an activity-silent state until the non-adjacent verb of the main clause triggers reactivation. Source-level analysis reveals that the reactivation first occurs in prefrontal regions followed by activation in the temporal cortex. This study provides crucial insights into the temporal and spatial dynamics of WM functions required for unification operations across embedded structures.

Keywords: natural language comprehension; working memory; WM reactivation; embedded language structures; MEG

Introduction

Recursion, as a fundamental computational property of human language (Hauser et al., 2002), manifests linguistically through embedded structures such as relative clauses. This remarkable ability presents a fundamental cognitive challenge, as it requires WM to maintain and integrate multiple elements across varying timescales during language comprehension. To understand this highly temporally dynamic process, this study employs MEG in combination with multivariate pattern analysis and source localization to capture the temporal dynamics and spatial distribution of neural activity during natural sentence processing. This study offers insights into how WM supports flexible maintenance and integration of linguistic information across embedded structures.

General Methods

Participants and experimental design

We recruited 34 native English speakers (20 females, 21.0±3.0 years old) with normal hearing and normal vision. Participants listened to English sentences with embedded structure (e.g. "The dog, who chases the cat, jumps over the mud.), followed by a visual figure. Then, they judged whether the figure's semantic content matched the sentence (Fig.1A). In the experiment we designed 48 unique sentences (including 4 different animals for subjects and 2 different verbs), each repeated 12 times, resulting in a total of 576 trials.

MEG recordings

Brain data was acquired with a 306-sensor TRIUX MEGIN system (204 orthogonal planar gradiometers and 102 magnetometers, Elekta, Finland). Visual stimuli were presented on a rear-projection screen. The pre-generated auditory stimuli were delivered by MEGIN Natus, consisting of non-magnetic transducers and air-conduction tubes. T1-weighted structural MRI images were acquired using a Simens 3.0T Prisma MRI system. The experimental program was controlled by MATLAB (MathWorks) and Psychtoolbox-3.

Source Localization

Source localization was performed according to the structural image data of each subject. First, the position of the individual brains in the MEG system was aligned with the MRI structural images (co-registration). Then, the MEG data for each trial were projected into the reconstructed cortical source space to obtain surface source estimates using the dynamic statistical parametric mapping method (dSPM) (Dale et al., 2000). The noise covariance matrix was calculated using the empty room signals on experimental days.

Neural decoding analysis

For the MEG decoding analysis, we employed a support vector machine (SVM) to classify and reveal the neural representation of the subjects ('dog'/'cat'/'fox'/'goat') during subject encoding phase (after subject ('dog') onset) and subject-verb integration phase (after Verb ('jumps') onset). A source-level analysis employing a representational similarity analysis (RSA) searchlight approach identified brain regions involved in WM encoding and reactivation during language comprehension.

Results

Neural representations of subject during WM encoding

To investigate the dynamic processes of subject encoding, we first aligned each MEG epoch according to the onset of the subject in each sentence and used an SVM to decode 'dog' vs 'cat' vs 'fox' vs 'goat' from 306 sensors. The decoding accuracy is significant after subject onset and returns to baseline level after 1.6s (Fig.1B, blue line cluster 1: 0s – 1.6s). This result demonstrates that subject categories can be successfully decoded from neural activity; however, the activation is transient. The neural representation gradually decays, then transitioning into an activity-silent state (Rose et al., 2016; Stokes, 2015; Wolff et al., 2017) until reactivated by strong syntactic cues (e.g. the verb, blue line cluster 2: 2.1s - 2.8s cluster $p < 0.05$, corrected).

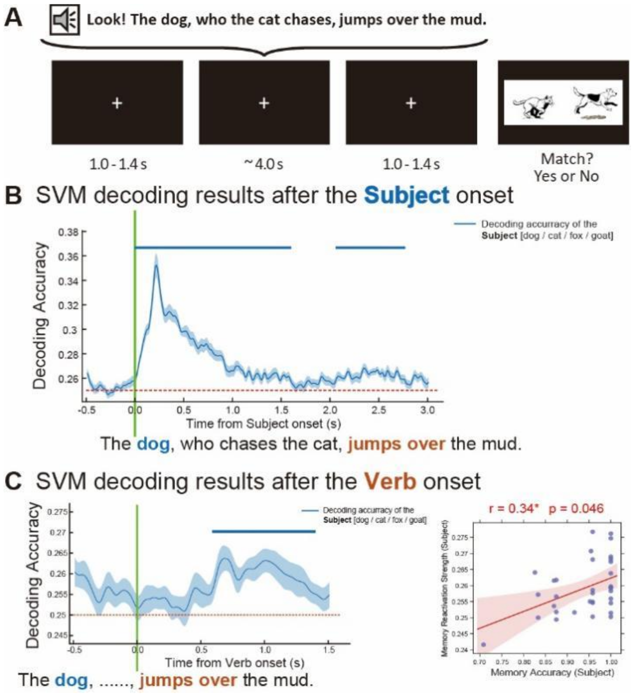

Fig. 1 Experimental paradigm and SVM decoding results during subject encoding and reactivation phases. (A) Experimental design for the MEG experiment (N = 34), illustrating the natural sentences with embedded structures and WM task. (B) Decoding accuracy of subject after subject onset. (C) Left: Decoding accuracy of subject after verb onset. Right: Correlation between subject reactivation strength and memory accuracy of subjects. Each dot indicates one participant (n=34, Spearman correlation). The shaded area represents the 95% CI.

WM reactivation of the earlier subject word when the verb is presented

To examine the relationship between subject (‘dog’) reactivation and the presentation of the verb (‘jumps’), we aligned MEG epochs to the verb onset. As shown in Fig. 1C, the representation-specific reactivation of the subject (‘dog’) occurred approximately 600ms after the verb (‘jumps’) onset, demonstrating that verbs trigger targeted reactivation of their syntactically related subjects from WM activity-silent state (Fig.1C left: blue line: 0.6s – 1.4s, cluster $p < 0.05$, corrected). The neural-behavioural correlation reveals a positive correlation between subject reactivation strength and memory accuracy of subjects (Fig.1C right: r=0.34, p=0.046; Spearman’s correlation). These results indicate that long-distance subject-verb integration involves selective reactivation of the subject, with the reactivation strength positively correlating with memory performance.

WM Reactivation in prefrontal cortex and temporal cortex

Source-level RSA searchlight analyses revealed the involvement of bilateral temporal cortex when the subject (‘dog’) is encoded into WM (peak at 0.22s). During WM reactivation, the representations of the subject ('dog') initially emerged in the prefrontal cortex (peak at 0.62s) and later engaged the bilateral superior temporal cortex (peak at 1.0s), suggesting that linguistic information is transformed to higher-order frontal regions, and triggers top-down reactivation in encoding related areas when integration is required.

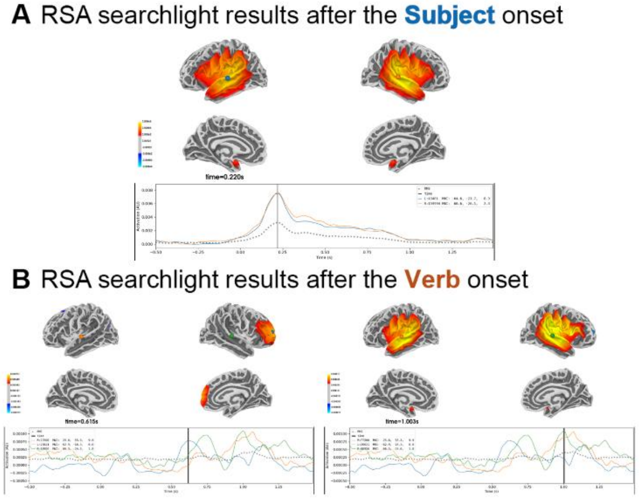

Fig. 2 Source-level searchlight results during subject encoding and reactivation periods. (A) WM encoding of subjects decoded from temporal lobes. (B) WM reactivation of subjects decoded from frontal and temporal lobes.

Conclusion

Our findings reveal representational-specific reactivation across embedded language structures. This suggests that WM for word information is maintained as activity-silent representations but can be flexibly retrieved when required later in the sentence. This reactivation signals initially emerges in prefrontal cortex before engaging temporal regions, pointing to a functional network supporting reactivation and unification operations (Buchsbaum, 2016; Hagoort, 2013; Pylkkänen, 2019). These dynamic processes illuminate how linguistic information is flexibly maintained and transformed in WM during natural language comprehension, offering crucial insights into the neural foundations of our uniquely human language abilities (Dijksterhuis et al., 2024; Lewis et al., 2006).

Acknowledgments

The work was supported by funding from a Wellcome Trust Discovery Award (grant number 227420 to OJ) and by the NIHR Oxford Health Biomedical Research Centre (NIHR203316 to OJ). The funders had no role in the preparation of the manuscript or decision to publish.

References

Buchsbaum, B. R. (2016). Working Memory and Language. In *Neurobiology of Language* (pp. 863–875). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00069-9

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. *Neuron*, *26*((1)), 55–67. https://doi.org/10.1016/S0896-6273(00)81138-1

Desbordes, T., King, J.-R., & Dehaene, S. (2024). Tracking the neural codes for words and phrases during semantic composition, working-memory storage, and retrieval. *Cell Reports*, *43*(3), 113847. https://doi.org/10.1016/j.celrep.2024.113847

Dijksterhuis, D. E., Self, M. W., Possel, J. K., Peters, J. C., Baaijen, J. C., Hanslmayr, S., Chelvarajah, R., Roux, F., Kolibius, L. D., Sawlani, V., Rollings, D. T., Dehaene, S., & Roelfsema, P. R. (2024). Pronouns reactivate conceptual representations in human hippocampal neurons. *Science*, *385*, 1478–1484. https://doi.org/science.org/doi/10.1126/science.adr2813

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00416

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, *366*(6461), 62–66. https://doi.org/10.1126/science.aax0050

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136–1139. https://doi.org/10.1126/science.aah7011

Stokes, M. G. (2015). “Activity-silent” working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546