Both semantic and form representations are pre-activated during sentence comprehension: Evidence from EEG Representational Similarity Analysis

Lin Wang (Tufts University, Harvard Medical School), Trevor Brothers (Tufts University, Harvard Medical School), Cheng Feng (Tufts University), Sophie Greene (Tufts University), Ole Jensen (University of Birmingham), Gina Kuperberg (Tufts University, Harvard Medical School)

It is well established that incoming words are facilitated in proportion to their predictability during language comprehension[1]. However, it remains unclear whether upcoming linguistic information is *pre-activated* before new bottom-up input becomes available, and if so, whether such pre-activation occurs at semantic and/or form levels of representation[2]. In the present study, we used Representational Similarity Analysis (RSA) in combination with EEG to address this question. The basic assumption of RSA is that unique representations are encoded as distinct spatial patterns of neural activity, and so representationally similar items (e.g. the same words) produce neural patterns that are more similar to each other than representationally distinct items (e.g. different words)[3]. By combining RSA with EEG, it is possible to determine *when* representationally specific information is encoded prior to the onset of incoming word[4]. In order to dissociate the time-course of form-based and meaning-based pre-activation, we capitalized on the ambiguity of homonyms — words that have the same orthographic and phonological form but distinct meanings (e.g. *bank*). Participants read highly constraining sentences that were predictive of either: (1) a homonym's subordinate meaning (e.g. a river *bank*), (2) its dominant meaning (e.g. a financial *bank*), or (3) a word that was semantically related to the dominant meaning (e.g. *loan*). Spatial RSA was conducted on EEG data at each time point prior to word onset to determine whether/when readers pre-activated *semantic* or *word-form* representations.

Design: We developed 84 triplets of highly constraining sentences (Table 1) (cloze: mean ± SD = 88% ± 8%). Each triplet contained a form-related homonym pair (*bank-bank*), with one member constraining for the homonym's subordinate meaning and the other constraining for its dominant meaning. Each triplet also contained a semantically related pair, with one member constraining for the homonym's dominant meaning and the other constraining for a word that was semantically related to this dominant meaning (*bank-loan*). In the EEG experiment, sentences in each triplet were presented in pseudorandom order and separated by at least 30 sentences. Each sentence was presented word by word (300ms per word + 400ms ISI). Participants (N=33) answered True/False comprehension questions following 1/6th of the sentences.

RSA Analysis: At each time point from -700ms before until the onset of critical words, we correlated spatial patterns of EEG activity (across 64 channels) *within* form related homonym pairs (e.g. *bank-bank*) and *within* semantically related pairs (e.g. *bank-loan*), and subtracted these values from the correlations produced *between* unrelated pairs (e.g. *bank-foot*, *bank-toes*, *loan-toes*) (Fig. 1). This difference reflects the increase in neural similarity associated with items with overlapping vs. non-overlapping representations. We then conducted cluster-based permutation tests (10,000 permutations) across the full prediction time window (-700ms to 0ms relative to critical word onset) to identify significant differences in spatial similarity across conditions.

Results: The semantically related pairs showed greater similarity effects (within-pairs > between-pairs) between -391ms and -309ms (p = .003) prior to the critical word onset, while the form related homonym pairs showed greater similarity between -53ms and -8ms (p = .025) (Fig. 2).

Discussion: These findings provide clear neural evidence for semantic and form *pre-activation* during the incremental comprehension of predictable sentences. Moreover, the earlier pre-activation of semantic than form information is consistent with a hierarchical generative framework[5], which posits that top-down pre-activation is propagated from higher to successively lower levels of the linguistic hierarchy over time.

Table 1. Examples of sentences

1a	The muddy sides of a river are called a bank.	Subordinate	Form related
1b	James went to deposit the check at his bank.	Dominant	
1c	To pay for college the student took out a loan.	Dominant-related	Between-pairs
2a	There are twelve inches in a foot.	Subordinate	
2b	He put a shoe on his left foot.	Dominant	Semantically related
2c	He had healthy nails on all his fingers and toes.	Dominant-related	

RSA methods

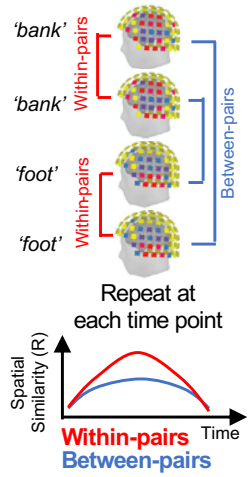

Fig.1. A schematic illustration of the spatial RSA analysis stream. First, for each trial, and at each time point, we extracted a vector of EEG data that represented the spatial pattern of activity across all 64 EEG channels. Second, we quantified the degree of spatial similarity of EEG activity produced by pairs of trials by correlating their spatial vectors. Third, we averaged the spatial similarity R-values separately for sentence pairs that predicted words with overlapping representations ("within-pairs") and for sentence pairs that predicted words without overlapping representations ("between-pairs"). Finally, we repeated this process at each time point, yielding time-series of R-values that reflected the degree of spatial similarity at each time sample between sentence pairs that predicted words with or without overlapping representations. The spatial similarity difference between the within-pairs and between-pairs reflected the increase in neural similarity associated with items with overlapping vs. non-overlapping representations.

RSA results

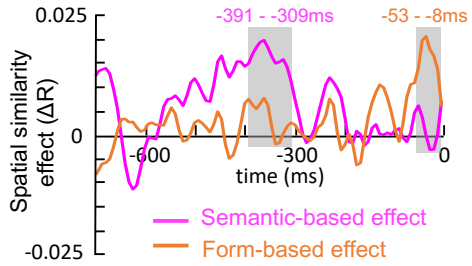

Fig. 2. Time course of semantic-based and form-based spatial similarity effects. The semantic-based effect was obtained by subtracting the spatial similarity/correlations within the semantically related pairs from the between-pair correlations, and the form-based effect was obtained by subtracting the spatial similarity/correlations within the form related homonym pairs from the between-pair correlations. Relative to the between-pairs, the spatial similarity was greater when the predicted words were semantically related (p = .003) between -391 and -309ms, and when the predicted words had the same word forms (p = .025) between -53ms and -8ms prior to the critical word onset.

References

[1] DeLong, Urbach, & Kutas. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121.

[2] Nieuwland. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, *96*, 367-400.

[3] Kriegeskorte, Mur, & Bandettini. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4.

[4] Wang, Kuperberg, & Jensen. (2018). Specific lexico-semantic predictions are associated with unique spatial and temporal patterns of neural activity. *ELife*, 7, e39061.

[5] Kuperberg, & Jaeger. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, 31(1), 32-59.