Title: Disentangling semantic composition and semantic association in the left temporal lobe

Abbreviated title: Semantic composition versus association

Authors:

Jixing Li (corresponding author), Post-doctoral associate, New York University Abu Dhabi

Saadiyat Island, Abu Dhabi, United Arab Emirates

+971 0561580376

jixingli@nyu.edu

Contributions: designed research, performed research, analyzed data, wrote the paper.

Liina Pylkkänen, Professor, New York University

70 Washington Square South, New York, NY 10012

New York University Abu Dhabi

Saadiyat Island, Abu Dhabi, United Arab Emirates

liina.pylkkanen@nyu.edu

Contributions: designed research, wrote the paper.

Number of pages: 38

Number of figures: 6 (excluding supplementary figures)

Number of words: Abstract: 192; Introduction: 636; Discussion: 1119

Conflict of interest: The authors declare no competing interests.

Acknowledgments: This work was supported by the NYU Abu Dhabi Institute under Grant G1001.

Abstract

Although composing two words into a complex representation (e.g., "coffee cake") is conceptually different from forming associations between a pair of words (e.g., "coffee, cake"), the brain regions supporting semantic composition have also been implicated for associative encoding. Here, we adopted a two-word magnetoencephalography (MEG) paradigm which varies compositionality ("French/Korean cheese" vs. "France/Korea cheese") and strength of association ("France/French cheese" vs. "Korea/Korean cheese") between the two words. We applied both univariate regression analyses and multivariate pattern classification to the source-localized MEG data in a bilateral language network. We show that the left anterior and middle temporal lobe (LMTL) are distinctively modulated by semantic composition and semantic association. Specifically, the LATL is mostly sensitive to high-association compositional phrases, while the LMTL responds more to low-association compositional phrases. Pattern-based directed connectivity analyses further revealed a continuous information flow from the anterior to the middle temporal region, suggesting that the integration of adjective and noun properties originated earlier in the LATL is consistently delivered to the LMTL when the complex meaning is newly encountered. Taken together, our findings shed light into a functional dissociation within the left temporal lobe for compositional and distributional semantic processing.

Significance Statement

Prior studies on semantic composition and associative encoding have been conducted independently within the subfields of language and memory, and they typically adopt similar two-word experimental paradigms. However, no direct comparison has been made on the neural substrates of the two processes. The current study relates the two streams of literature, and appeals

to audiences in both subfields within cognitive neuroscience. Disentangling the neural computations for semantic composition and association also offers insight into modeling compositional and distributional semantics, which has been the subject of much discussion in natural language processing and cognitive science.

Introduction

When we hear a pair of words such as "coffee" and "cake", we could form a complex meaning of a coffee-flavored cake, or recall an experience where coffee and cake occurred together. These two processes, though conceptually different, have been localized to similar brain regions in the left temporal lobe. For example, a minimal adjective-noun phrase such as "red boat" elicits increased activity in the left anterior temporal lobe (LATL) compared to non-compositional word lists such as "cup, boat" (Bemis and Pylkkänen, 2011, 2013), and a similar effect has been observed for a language with the reverse word order (Westerlund et al., 2015) and for American Sign Language[4], suggesting a role of the LATL in conceptual combination (Blanco-Elorrieta et al., 2018). However, the LATL is also activated when forming arbitrary associations between pairs of words such as "ring" and "cheese" (Jang et al., 2017), and shows greater oscillatory responses for word pairs with stronger associations (Teige et al., 2018, 2019). Additionally, a temporary virtual lesion induced by repetitive transcranial magnetic stimulation (rTMS) over the LATL significantly slows synonym judgment times (Pobric et al., 2007), mirroring a core feature of semantic dementia (SD) patients with a focal atrophy of the ATL (Galton et al., 2001; Hodges et al., 1992; Nestor et al., 2006). The LATL is also engaged in human face-name (Wang et al., 2017) and face-location (Nieuwenhuis et al., 2011) associations, and single neuron recordings in the ATL of macaque monkey's brain suggested that a group of pair-encoding neurons responded only to associated pairs

of abstract shapes during the training phase (Hirabayashi et al., 2014; Sakai and Miyashita, 1991).

Prior studies on semantic composition and associative encoding have been conducted independently within the subfields of language and memory, and they typically adopt similar two-word experimental paradigms. However, no direct comparison has been made on the neural substrates of the two processes. Here we adopted the same two-word paradigm but crossed compositionality and strength of association between the two words using the adjectival and noun forms of a country word and our world knowledge of food. “France cheese” and “French cheese”, for example, are highly-associated country-food pairs compared to “Korea cheese” and “Korean cheese”. But both “French cheese” and “Korean cheese” are combinatory adjective-noun phrases whereas “France cheese” and “Korea cheese” are merely word lists. This 2 x 2 design enabled us to separate the effects of association and composition during two-word comprehension (Figure 1).

We collected 42 native English speakers’ MEG data while they viewed the words one by one on a screen. Highly time-resolved MEG data allow us to characterize the temporal dynamics of semantic association and semantic composition in addition to their spatial locations. Our relatively large sample size also provides more statistical power compared to many prior neuroimaging studies. We coded association and composition as binary variables and applied a mass univariate regression analysis to source-localized MEG data within a bilateral language mask. We then performed a multivariate pattern classification analysis within the significant clusters identified by the regression analysis. To further understand the information flow between the active brain regions, we conducted a directed connectivity analysis (Goddard et al., 2016; Lyu et al., 2019) using the representational dissimilarity matrices (RDMs) of the MEG data within the functional regions of interest (fROIs).

With this combination of methods, we show that the LATL is most sensitive to the contrast between high and low associative compositional phrases, while the LMTG is mainly driven by the distinction between low-association compositional phrases and low-association lists. Directed connectivity analyses suggest a continuous information flow from the LATL to the LMTG. Taken together, our results provide novel evidence that the LATL and the LMTL are distinctly modulated by semantic composition and semantic association. We postulate a functional dissociation within the left temporal lobe for compositional and distributional semantic processing.

Materials and Methods

Experiments and participants. We conducted two variates of experiment, both employed a 2 x 2 design with associative strength (Low, High) and compositionality (Comp, List) as factors. In the first experiment, there are 60 trials for each of the 4 conditions and 60 single word controls consisted of 5 "x"s and the food noun (e.g., "xxxxx cheese"), forming a total of 300 unique trials. In the second experiment, each condition contains 45 trials and 45 single word controls, forming a total of 360 unique trials. The single word trials consisted of length-matched consonant strings and the food noun (e.g., "xjtgbv cheese").

Participants were healthy young adults with normal hearing and normal or corrected-to-normal vison. All strictly qualified as right-handed on the Edinburgh handedness inventory (Oldfield, 1971). They self-identified as native English speakers and gave their written informed consent prior to participation, in accordance with New York University and New York University Abu Dhabi IRB guidelines. 30 volunteers participated in Experiment 1, 9 of them were removed from data analysis: 8 due to excessive movement or drowsiness during MEG recording, and 1 due to bad performance on the behavioral task (less than 75% accuracy). 26 volunteers participated in

Experiment 2, 5 of them were removed from further data analysis: 4 due to excessive movement or drowsiness during MEG recording, and 1 due to bad performance on the behavioral task. Therefore, a total of 42 participants (24 female, mean age=22.3 years, SD=6.3) were included in the analyses. No participant attended both experiments.

Experiment 1 was conducted at the Neuroscience of Language Lab at New York University Abu Dhabi, and Experiment 2 was conducted across research facilities of the Neuroscience of Language Lab at New York University and New York University Abu Dhabi. 8 participants' data were collected at the New York facility.

Stimuli. The stimuli comprised of a country adjective/noun and a food noun, with strength of association (Low, High) and compositionality (Comp, List) as factors (Figure 1A). We quantified the strength of association between the country and food word using the cosine similarity score between the GloVe embeddings[17] of the two words: High associative phrases (e.g., "French/France cheese") have a cosine similarity greater than 0.3 and low associative pairs (e.g., "Korean/Korea cheese") have a cosine similarity lower than 0.15 (Supplementary Table 1). Note that, although the noun-noun lists (e.g., "France cheese") were not syntactically composable, participants may still attempt conceptual composition during the task. To ensure that the list conditions were indeed non-compositional as intended, we included a single-word control condition. The single-word condition consisted of either a "xxxxx" or a length-matched consonant string (e.g., "xjtgbv") and the critical noun word (i.e., "cheese"). Half of the participants (n=21) saw the "xxxxx" while the other half saw the consonant strings. One ancillary aim of our study was to examine the impact of these two different types of visual baselines, but they patterned the same as regards our main manipulation (see Supplementary Figure 1)

Experiment procedures. Before recording, each subject's head shape was digitized using a Polhemus dual source handheld FastSCAN laser scanner (Polhemus, VT, USA). Participants then completed the experiment while lying supine in a dimly lit, magnetically shielded room. At the Abu Dhabi facility, MEG data were recorded continuously using a whole-head 208 channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan); at the New York facility, MEG data were collected using a whole-head 156-channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan).

Participants completed the experiment lying supine in a dimly lit magnetically shielded room (MSR). Stimuli were projected using PsychoPy2 (Peirce, 2019) onto a screen approximately 80 cm away from the participants' eyes. The two words were presented for 300 ms each, in white 30-point Courier New font, on a gray background. An image appeared on screen after the target words and remained until subjects indicated whether it was a Match or a Mismatch to the preceding words, by pressing a button with either the index or middle finger of their left hand. No feedback was provided. A blank screen was presented for 300 ms between each word and image. The inter-trial interval was normally distributed with a mean of 400 ms (SD=100ms). Order of stimulus presentation was pseudo-randomized such that half of the trials were a Match and the other half a Mismatch. Each participant received a unique randomisation. Images for the categorization task were chosen to encourage concentration on both country nouns/adjectives and food nouns. For the two-word conditions, the target picture shows a food item with a country flag, and both the country and the food words were required to match; for the one-word condition, the picture contains only a food item. All images were color photographs found on Google Images. The whole recording session, including preparation time and practice, lasted around 40 minutes.

MEG data acquisition and pre-processing. MEG data were recorded continuously at a sampling rate of 1000 Hz with an online 0.1 to 200 Hz band-pass filter. The raw data were first noise reduced via the Continuously Adjusted Least-Squares Method (Adachi et al., 2001) and low-pass filtered at 40 Hz. Independent component analysis (ICA) was then applied to remove artifacts such as eye blinks, heart beats, movements, and well-characterized external noise sources. The MEG data were then segmented into epochs spanning 100 ms pre-stimulus onset to 1200 ms following, and baseline corrected relative to the initial 100 ms pre-stimulus interval. Epochs containing amplitudes greater than an absolute threshold of 2000 fT were automatically removed. Additional artefact rejection was performed through manual inspection of the data, removing trials that were contaminated with movement artefacts or extraneous noise. The whole epoch rejection procedure results in an average rejection rate of 9.4% (SD=6.5%) for each participant.

Cortically constrained minimum-norm estimates (Hämäläinen and Ilmoniemi, 1994) were computed for each epoch for each participant. To perform source localization, the location of the participant's head was coregistered with respect to the sensor array in the MEG helmet. For participants with anatomical MRI scans (n=5), this involved rotating and translating the digital scan to minimize the distance between the fiducial points of the MRI and the head scan. For participants without anatomical scans, the FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) "fsaverage" brain was used, which involved first rotation and translation and then scaling the average brain to match the size of the head scan. A source space of 2562 source points per hemisphere was generated on the cortical surface for each participant. The Boundary Element Method (BEM) was employed to compute a forward solution, explaining the contribution of activity at each source to the magnetic flux at the sensors. Channel-noise covariance was estimated

based on the 100 ms intervals prior to each artifact-free trial. The inverse solution was computed from the forward solution and the grand average activity across conditions. To lift the restriction on the orientation of the dipoles, the inverse solution was computed with "free" orientation, meaning that the inverse operator places three orthogonal dipoles at each location defined by the source space. When computing the source estimate, only activity from the dipoles perpendicular to the cortex were included. For each trial within one condition, the same inverse operator for that condition was applied to yield dynamic statistical parameter maps (dSPM) units (Dale et al., 1999) using an SNR value of 3. All data preprocessing steps were performed with MNE-python (v.0.19.2; Gramfort, et al., 2014) and Eelbrain (v.0.25.2; Brodbeck et al., 2014).

Quantifying semantic associations. Semantic vectors for the country and food words were derived using the well-known GloVe word embeddings model (Pennington et al., 2014; freely available at https://nlp.stanford.edu/projects/glove/) trained on Common Crawl (https://commoncrawl.org/), which contains petabytes of raw web page data, metadata extracts and text extracts. GloVe model embodies the "distributional hypothesis" that words with similar meaning occur in similar contexts in an artificial neural network approach. Having obtained a 300-dimensional vector for each word, we then quantified how semantically associated each country word was to the food word by calculating the cosine similarity score between the two word vectors. Overall, our approach generated high associative phrases (e.g., "French/France cheese") with a cosine similarity score greater than 0.3 and low associative pairs (e.g., "Korean/Korea cheese") with a cosine similarity score lower than 0.15. Figure 1B visualizes all the word embeddings on a 2D scatter using t-SNE (van der Maaten, 2014).

Behavioral data analysis. We applied linear mixed effect model to the accuracy and reaction times (RTs) of participants' responses to the picture-matching task, with fixed effects for association and composition (with the interaction term), and a random intercept for subjects. To control for the word frequency effect, we also included the frequency of the first and second word as fixed effects:

$$\text{RT/accuracy} = \text{word1_frequency} + \text{word2_frequency} + \text{association} \times \text{composition} + (1|\text{subject})$$

Word frequency was calculated based on the combined word counts from the Google Book unigrams (http://books.google.com/ngrams) and the SUBTLEXus corpus (Brysbaert and New, 2009). RTs above or below than 3 standard deviations were removed under the assumption that these represented errors or distractions rather than task-related responses. Word frequencies and RTs were log-transformed. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. *p*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.

Mass univariate multiple regression. We coded the composition and association factors as binary variables and performed a two-stage multiple regression with single-trial source estimates as dependent variables, and compositionality, strength of association and their interaction as predictors (Figure 3A). The one-word trials were excluded from the regression analysis. Given the significant effects of word frequency on the behavioral results, we also included frequency of the first and second word as control variables. At the first stage, we applied an ordinary least squares regression for each subject's single-trial source estimates for each source within a bilateral language mask at each millisecond from 100 millisecond before the onset of the first word to 1200 millisecond post-stimulus onset. The language mask (see the light pink region in Figure 4A and 4E) covered regions including the whole temporal lobe, the inferior frontal gyrus (IFG; defined as

the combination of BAs 44 and 45), the ventromedial prefrontal cortex (vmPFC; defined as BA11), the angular gyrus (AG; defined as BA39) and the supramarginal gyrus (SMA; defined as BA 40). The left AG and vmPFC have also been implicated in previous literature on conceptual combination (Bemis and Pylkkänen, 2011; Price et al., 2015), and the LIFG and the LMTG have been suggested to underlie syntactic combination (Flick and Pylkkänen, 2020; Haggort, 2005; Lyu et al., 2019; Matchin et al., 2019; Matchin and Hicock, 2020). The medial temporal lobe, on the other hand, has been held critical for associative memory in addition to the anterior temporal lobe (e.g., Day et al., 2003; Ison et al., 2015; Miller et al., 2013; Rey et al., 2018). As a control, we also included these regions in the right hemisphere. The first-stage regression resulted in a β coefficient for each variable at each source and each timepoint for each subject.

At the second stage, we performed a one-sample *t*-test on the distribution of the β values across subjects for each variable separately, again at each source and each timepoint, to test if their values were significantly different from zero. The *t*-tests are one-tailed such that *t*-values with the same polarity are clustered together as separate regions. We searched for significant spatiotemporal clusters within the broadly defined bilateral language mask instead of simply the LATL in order to maintain maximal ability to identify potentially distinct localizations for association and composition. Clusters were formed based on the *t*-values that were contiguously significant through time and space, at a level of $p < .05$. Only clusters that contained a minimum of 10 sources and spanned at least 10 ms were entered into a cluster-based spatiotemporal permutation test (Maris and Oostenveld, 2007). This involved randomly shuffling 0 and the β coefficient for each participant, repeating the mass univariate one-sample *t*-test and cluster formation within the analysis time window of 0-1200 ms after the sitmulis onset. This procedure was performed 10,000 times, resulting in a distribution of 10,000 cluster-level statistics. Each of

the observed clusters was subsequently assigned a *p*-value based on the proportion of random partitions that resulted in a larger test statistic than the observed one. The center of mass for each spatiotemporal cluster in Montreal Neurological Institute (MNI) coordinates was calculated by weighting the position of each constituent source by its mean *t*-value over the time window of the cluster.

Multivariate pattern classification. To investigate more fine-grained encodings of the four conditions in the left anterior and middle temporal lobe, we conducted a multivariate pattern classification analysis within the LATL and LMTL fROIs identified by the univariate multiple regression analysis (Figure 3B). We selected the most responsive sources in the two fROIs at a threshold of $|t| > 1$. This resulted in an LATL fROI with 161 sources, and an LMTL fROI with 108 sources. Under the assumption that patterns of brain activation contain information that distinguishes between the experimental conditions, we applied classification to 4 pairwise combinations of our 4 conditions: high-association phrases vs high-association lists, low-association phrases vs low-association lists, high-association phrases vs low-association phrases, and low-association phrases vs low-association lists. The high-association phrases vs high-association lists classification and the low-association phrases vs low-association lists classification compares composition across the association levels, and the high-association phrases vs low-association phrases classification and the low-association phrases vs low-association lists classification compares association across the composition levels.

The general analysis pipeline comprised of preprocessing, data partitioning and classification steps. In preprocessing, we first down-sampled the single trial source estimates by averaging across 5 consecutive milliseconds bins. We then combined the trials in each condition

to 15 pseudo-trials by randomly dividing the trials into 15 partitions and averaging across the trials within partitions. These two steps were performed to decrease the computational costs and increase the signal-to-noise ratio. The resulting data pattern for one pseudo trial in the LATL fROI is a 161 (number of sources) x 206 (number of down-sampled timepoints) matrix, and the data for one pseudo-trial in the LMTL fROI is a 108 x 206 matrix. We further applied a multivariate noise normalization (Guggenmos et al., 2018) to the pseudo trials, where the data pattern x was normalized by means of a covariance matrix Σ: $x^* = \Sigma^{-\frac{1}{2}} x$. The covariance matrix Σ was computed on the full time points of the data.

After preprocessing, we trained 4 linear Support Vector Machine (SVM) classifiers, each applied to a pairwise combination of the 4 conditions. The SVM is a widely used classification method due to its favorable characteristics of high accuracy, ability to deal with high-dimensional data and versatility in modeling diverse sources of data (Schölkopf et al., 2004). Linear SVM classifier was chosen as it has been shown to out-perform other classifiers on both MEG (Guggenmos et al., 2018) and fMRI data (Misaki et al., 2010). We performed a leave-one-stimulus-pair-out cross-validation procedure, where one randomly selected pseudo-trial was taken from each condition as a test sample and all remaining stimuli were used for classifier training. 100 randomized assignments of trials to pseudo-trials and cross-validation folds were applied with the randomly selected left-out pairs of pseudo-trials. Classifier performance was estimated by averaging across the 100 permutations. The binary classifiers were separately applied to all timepoints. The regularization parameter C of the classifiers was selected by grid search in the range of $C = 2^{-5}$ to 2^{15} using five-fold cross-validation within the training data.

The same analysis pipeline was applied to each subject. Classification accuracy averaged over subjects at each timepoint minus the chance level of 50% was submitted to a one-sample one-

tailed *t*-test with 10,000 permutations. The permutation test window was between 600-1200 ms (Figure 3B). We used the python scikit-learn package (0.22.1) for the SVM analyses.

Directed connectivity analysis. To further understand how LATL and LMTL exchange information during the two-word comprehension process, we conducted a directed connectivity analysis to estimate the temporal pattern of information flow between the two regions. The underlying logic is similar to that of the Granger causality analysis (Granger, 1969), namely, if the activity in region A has a causal effect on the activity in region B, then activity at a past time point of region A should contain information that helps predict current activity in region B above and beyond the previous activity in region B alone. Here we performed a multivariate pattern-based connectivity analysis which has been applied previously to estimate information flows between brain regions on both visual (Goddard et al., 2016) and language processing (Lyu et al., 2019).

We first calculated the data representational dissimilarity matrices (RDMs) of the two fROIs at time point $D(A, t)$ and $D(B, t)$ as 1 minus the pairwise Pearson *r*'s correlation among the conditions. These RDMs indicate the degree to which different levels of associative strength and compositionality evoke similar or distinct response patterns in the neural population. To increase the signal-to-noise ratio, we adopted the same preprocessing procedures in the MVPA analysis, including downsampling the source estimates by averaging across 5 consecutive milliseconds bins and combining the trials in each condition to 15 pseudo-trials. We then quantified the directed activity from region A to region B as the partial correlation coefficient between $D(A, t\text{-}dt)$ and $D(B, t)$, partialling out $D(B, t-dt)$, where dt is the time interval between the current time point and the previous time point. To avoid bias due to the choice of directionality and any specific previous time point, we computed the directed connectivity from region A to region B and from region B to region A for each time points from the onset of the second word to the following 600 ms, with

a *dt* value of 5 ms before the current timepoint up till 600 ms (i.e., 5, 10, 15, …600 ms before the current time point). The extended time dimension allows us to determine the extent to which the current activity in the target region is correlated with the source region's activity at each time point within the previous 600 ms. The same procedure was applied to each subject to obtain an average partial correlation coefficient matrix (Figure 3C).

We report the results of this analysis in terms of the difference between the two directions of the information flow (A→B – B→A). Significance of the difference was determined by a one-sample *t*-test with 10,000 permutations.

Data and code availability. Experimental stimuli and MEG single-trial source estimates for each participant is available to download via Open Science Framework (OSF) at https://osf.io/ea4s6/. All analyses were performed using custom codes written in Python, making heavy use of mne, eelbrain and scikit-learn. The analysis codes can be downloaded at https://osf.io/ea4s6/.

Results

Behavioral results. Overall, participants achieved a high accuracy of 90.9% (SD=28.2%) with a mean RT of 1.07 s (SD=0.77 s). The mixed-effects regression analysis revealed a significant effect of association for both accuracy and reaction time, where higher association between the two words increased accuracy (t=1.97, p=0.049) and reduced reaction time (t=-2.16, p=0.03). However, the composition factor was not significant for either accuracy or reaction time (Figure 2A). We also observed a highly significant word frequency effect on both accuracy and RTs, such that more frequent first word reduced reaction time (t=-4.68, p~=0), and more frequent second word both increased accuracy (t=3.83, p~=0) and reduced reaction time (t=-6.55, p~=0). This

typical slowdown in responses for less frequent words indicates that the stimuli are being perceived as intended (Figure 2B).

Spatiotemporal clusters for association and composition. We identified one significant cluster of 302 sources for the interaction effect between association and composition from 544 to 846 ms (p=0.019), which is 54 ms before the onset of the second word till 246 ms afterwards. The cluster covered regions including the anterior temporal lobe, the anterior parts of the inferior, middle and superior temporal gyrus (Figure 4A). The direction of the interaction is positive, which means that high-association compositional phrases elicited higher activity compared to other conditions (Figure 4B). We then calculated the mean fitted activation for the four conditions after controlling for the word frequency effect. The results showed two peaks of opposite polarity within this cluster: For the compositional phrases, high-association was correlated with higher positive activity at the first positive activation period from 544-750 ms, and lower negative activity at the second negative activation period from 750-846 ms. For the non-compositional list conditions, however, high-association was correlated with less positive activity than low-association at the first positive activation period, and high-association was correlated with increased negative activity than low-association at the second negative activation period (Figure 4C,D). The peak activity of this interaction effect was localized to the left anterior temporal lobe (MNI coordinates: [-54,5,-11]) at around 680 ms, which is 80 ms after the onset of the second word.

The effect of composition was associated with one significant cluster of 412 sources from 675 to 1050 ms (p=0.032) after the onset of the first word (i.e., 75-450ms after the onset of the second word). The cluster covered regions including the middle to posterior parts of the inferior, middle and superior temporal gyrus (Figure 4E). The sign of the β coefficient is negative, such that

compositional phrase induced more negative activity (Figure 4F). The fitted mean activation after controlling for the word frequency effects showed one positive and one negative activation period. At the first positive period at around 675-750 ms, there is increased positive activity for the list conditions; at the second negative activation period, there is increased negative activity for the composition conditions (Figure 4G,H). However, as shown in Figure 4G, this effect of composition seems to be mainly driven by the difference between composition and lists within the low-association level. The peak activity of this effect was localized to the left middle temporal gyrus (MNI coordinates: [-49,-11,-19]) at around 815 ms, which is 215 ms after the onset of the noun word, consistent with the typical time window for semantic composition (see Pylkkänen 2019 for a review). No significant cluster was found for any of the effects in the right hemisphere within the mask, suggesting a highly left-lateralized neural activity for association and composition.

We also observed a large significant cluster of 498 sources for lexical frequency of the second word in a region spanning left medial-posterior temporal cortex from 820-1200 ms (p=0.038), that is, 220-600 ms after the onset of the second word. Here, higher word frequency decreased activation, consistent with most findings on the effect of frequency (Embick et al., 2001; Simon et al., 2017). The peak activity is located at the inferior middle temporal gyrus (MNI coordinates: [-57,-33,-16]) at around 926 ms. In contrast to word two frequency, word one frequency was not correlated with any significant cluster within the left language mask.

To confirm that the list conditions were indeed perceived as non-compositional as the single-word trials, we further conducted post-hoc pair-wise t-tests within the LATL and the LMTL clusters. We expected that the activation timecourses of the list conditions pattern with that of the single-word condition and this prediction was borne out: No significant temporal clusters were observed for the contrast between the two list conditions and the single word condition in either

fROI. On the contrary, the two composition conditions were significantly different from the single-word condition in both the fROIs (see Supplementary Figure 1).

Multivariate pattern classification within the active brain regions. The classification results are shown in Figure 5. In the LATL functional ROI, classification accuracy for high-association composition vs. low-association composition was significantly higher than chance from 850-935 ms (p=0.019). A cluster for the low-association composition vs. low-association lists classifier was identified from 895-950 ms, but it did not survive the permutation significance test (p=0.12). Decoding performance for other classifiers was not significantly higher than chance. Compared with the multiple regression results, we observed the same distinction between high-association and low-association compositional phrases. However, high-association and low-association lists cannot be successfully decoded from each other, although their averaged activation differ. This suggests a smaller effect size for the distinction between the list conditions in the LATL.

In the LMTL functional ROI, the classifier for the low-association composition vs. low-association lists was significant at around 830-935 ms (p=0.035). Other classifiers did not perform significantly better than chance. This is consistent with the pattern observed in the regression results, where the effect of composition seems to be mainly driven by the distinction between composition and list at the low-association level.

Information flow between the active brain regions. So far we have demarcated the LATL and the LMTL for high- and low-association compositional phrases, respectively. However, in the temporal dimension, the time windows for the two effects are roughly the same from both the regression and the classification analyses. With the pattern-based directed connectivity analyses,

we are able to disentangle the two effects at the temporal dimension. Comparison between the connectivity measures of the two directions revealed significant partial correlations from LATL to LMTL at around 0-300 ms after the onset of the second word. The significant time interval before the current timepoint ranged from almost all the time intervals between 5 to 600 ms (Figure 6A). No significant partial correlations from LMTL to LATL was found by the contrast measure (Figure 6B). This suggests that information about associative strength and compositionality of the two words generated in the LATL within the previous 600 ms is continuously delivered to the adjacent LMTL from the onset of the second word to 300 ms post-stimulus onset.

Although the LATL and LMTL functional ROIs in our directed connectivity analysis are selected based on the regression results, this method is still largely data-driven as the connectivity strength is quantified by the partial correlations between the data RDMs. To investigate whether the connectivity results were specific to semantic association and composition or simply reflect intrinsic interactions between two adjacent regions, we conducted a supplementary control analysis with the corresponding brain regions in the right hemisphere. Since the RATL and RMTL were not significant to either association or composition effect in the regression analysis, the patterning of data RDMs in the two regions should bear no relation to our stimulus manipulation. Therefore, we expect different connectivity profiles between the two regions in the right hemisphere.

The results supported our prediction: although we did find directed connectivity from RATL to RMTL, the temporal dynamics of the information flow were quite different from that between the LATL and the LMTL. Supplementary Figure 2A showed the contrast between the directed connectivity measures from RATL to RMTL and from RMTL to RATL. There are significant partial correlations from RATL to RMTL from around 150 to 250 ms after the onset of

the second word, with a delay of 300-450 ms. No significant partial correlations were found from RMTL to RATL by the contrast measure (Supplementary Figure 1B). Compared to the RATL and RMTL connectivity effect, the information flow from LATL to LMTL is more sustained with a much longer delay window. These additional effects cannot be solely attributed to intrinsic connectivity between adjacent brain regions.

Discussion

Prior work has identified similar brain regions for both semantic composition and associative encoding. This raises the question of whether semantic composition is simply a form of verbal association. In theory, semantic compositionality allows human language to express an infinite number of complex expressions from their constituent meanings (Frege, 1980). This productive combinatory process is not dependent on the semantic relatedness of the constituent meanings. It engages even for totally novel combinations, such as "green idea". In practice, however, words that occur together are usually semantically related. As stated in distributional semantics, "a word is characterized by the company it keeps" (Firth, 1957), thus a common complex expression like "green grass" involves both combinatory and associative processes.

By orthogonalizing these two factors in the experimental design, we show distinct profiles in the LATL and the LMTL: In the LATL, we observed higher activation for high-association compositional phrases such as "French cheese" than for low-association compositional phrases such as "Korean cheese". Multivariate pattern classification also shows distinctive activity patterns between the two conditions. These results suggest that the LATL is most sensitive to commonly encountered complex expressions in everyday life such as "green grass". In the LMTL, the regression analysis and the multivariate pattern classification results indicated that this region is

mainly sensitive to the distinction between the composition and list conditions at the low-association level, i.e., the distinction between “Korean cheese” and “Korea cheese”. In other words, the LMTL responds more strongly to newly encountered compositional phrases such as “green idea”.

We set out to disentangle association and composition in the LATL and we replicated previous findings that the LATL is essential for both factors. In particular, we provide novel evidence that the LATL responded most strongly when the association and composition factors are combined. In addition to the LATL results, we also found an LMTL activity to low-association compositional phrases. Compared with high-association compositional phrases, low-association compositional phrases form a newly-encountered meaning, hence likely involve more syntactic processing for a successful parsing. There is ample evidence that the LMTL supports syntactic combination (Flick and Pylkkänen, 2020; Lyu et al., 2019; Matchin et al., 2019; Matchin and Hicock, 2020). For example, the left pMTG in verb-noun combination shows the strongest model fit in the left middle temporal gyrus (Lyu et al., 2019), and both noun and verb phrases (e.g., “the frightened boy” and “frightened the boy”) inducing greater activity in the superior temporal sulcus compared to word lists (Matchin et al., 2019). Moreover, post-nominal adjective-noun composition correlates with activity in the left medial to posterior temporal lobe when semantics is held constant (Flick and Pylkkänen, 2020).

Our findings also align well with the controlled semantic cognition (CSC) framework (Lambon Ralph et al., 2017), which proposes a functional dissociation of the anterior and middle temporal regions as well. Under this framework, the ATL integrates information from different sources into a coherent concept, whereas the posterior middle temporal gyrus (pMTG) retrieves and manipulates semantic knowledge. Evidence for this functional dissociation comes from studies

on semantic disorders showing different patterns of impairment: patients with focal ATL atrophy exhibit symptoms of semantic dementia (SD), with performance strongly influenced by concept familiarity (Ding et al., 2020). In contrast, patients with temporoparietal damage are usually diagnosed with semantic aphasia (SA), characterized by their deficits in controlled semantic retrieval (Hoffman et al., 2018), such as retrieving infrequent semantic associations. Mirroring the two patients' profiles, we also observed the LATL activation for semantically related word pairs, and the LMTL response to infrequent semantic associations. We additionally demonstrate that compositionality interacts with semantic association, supporting a crucial role of syntactic knowledge in semantic cognition, and vice versa.

At the temporal dimension, the regression analysis showed a slightly earlier LATL activity compared to the LMTL activity with a largely overlapping time window from 650-800 ms, replicating previous findings that composition occurs around 180-250 ms after the onset of the second word (see Pylkkänen for a review). The time periods for significant decoding performance within the two regions were a bit later at around 850-930 ms period. Crucially, the LATL and LMTL activity did not differ much in the time domain across the regression and decoding analysis. The pattern-based directed analysis, however, revealed that activity originated in the LATL 600 ms before it was continuously delivered to the LMTL within 300 ms after the stimulus onset, suggesting a prominent information flow from the LATL to the LMTL. The LATL and the LMTL have been shown to be functionally connected during semantic association judgements (Jackson et al., 2016), and the LATL was assumed to connect to the LMTL among other modality-specific regions under the "hub-and-spoke" model of semantic cognition (Patterson et al., 2007). Our results provide novel evidence on the direction of the functional connection between the LATL and the LMTL. Our regression and decoding analyses showed that the LATL is more sensitive to

the high-association compositional phrases while the LMTL is responding to the low-association compositional phrases. The information flow from the LATL to the LMTL further suggests that high-association composition precedes low-association composition. We interpreted these results as reflecting a functional dissociation where the LATL supports "shallow" meaning composition of common phrases, while the LMTL is engaged to interpret novel phrases via syntactic composition. At the temporal dimension, "shallow" meaning formation occurs earlier in the LATL, and triggers syntactic composition in the LMTL when the complex meaning is unfamiliar.

The interaction of composition and association in the left temporal lobe offers insight into the relation between compositional and distributional semantics, which has been the subject of much discussion in natural language processing and cognitive science. Distributional semantic models work by building co-occurrence vectors for every word in a corpus based on its context. They provide concrete information of word meaning but do not scale up to larger constituents of text, such as phrases or sentences. Compositional formal semantics complements this approach and various models have been proposed to combine the two paradigms (e.g., Baroni et al., 2014; Mitchell and Lapata, 2008). These models typically contain one uniform operator such as addition or multiplication for all types of compositional phrases. Here we showed that the LATL and LMTL respond differently to high- and low-association compositional phrases, suggesting potentially different neural mechanisms for combining constituents with different associative strength.

In summary, we show evidence for distinct profiles in the left temporal lobe where the LATL is mostly sensitive to high-association compositional phrases, while the LMTL responds more to low-association compositional phrases. We also observed a prominent information flow from the LATL to the LMTL, suggesting that the integration of adjective and noun properties

originated earlier in the LATL is consistently delivered to the LMTL when the complex meaning is newly encountered.

References

Adachi Y, Shimogawara M, Higuchi M, Haruta Y, Ochiai M (2001) Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE T Appl Supercon 11: 669-672.

Baroni M, Bernardi R, Zamparelli R (2014) Frege in space: A program for compositional distributional semantics. Ling Lang Tech 9:241-346.

Bemis D, Pylkkänen L (2011) Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistics phrases. J Neurosci 3:2801-2814.

Bemis D, Pylkkänen L (2013) Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. PLoS ONE 8:e73949.

Blanco-Elorrieta E, Kastner I, Emmorey K, Pylkkänen L (2018) Shared neural correlates for building phrases in signed and spoken language. Sci Rep 8:5492.

Brodbeck C et al (2020) Eelbrain-v032 DOI:105281/zenodo3923991.

Brysbaert M, New B (2009) Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behav Res Methods 41:977-990.

Dale AM, Liu A, Fischl B, Buckner R, Belliveau J, Lewine J, Halgren E (2000) Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26:55– 67.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis I Segmentation and surface reconstruction. NeuroImage 9:179-194.

Day M, Langston R, Morris RG (2003) Glutamate-receptor-mediated encoding and retrieval of paired-associate learning. Nature 424:205–209.

De Falco ED, Ison MJ, Fried I, Quian Quiroga R (2020) Long-term coding of personal and universal associations underlying the memory web in the human brain. Nat Commun 7:13408.

Ding J, Chen K, Liu H, Huang L, Chen Y, Lv Y, Yang Q, Guo Q, Han Z, Lambon Ralph MA (2020) A unified neurocognitive model of semantics, language social behavior and face recognition in semantic dementia. Nat Commun 11:2595.

Embick D, Hackl M, Schaeffer J, Kelepir M, Marantz A (2001) A magnetoencephalographic component whose latency reflects lexical frequency. Cognitive Brain Res 10:345-348.

Firth J (1957) A synopsis of linguistic theory, 1930-1955. In: Studies in linguistic analysis, pp1-32. Oxford: Philosophical Society.

Flick G, Pylkkänen L (2020) Isolating syntax in natural language: MEG evidence for an early contribution of left posterior temporal cortex. Cortex 127:42-57.

Frege G (1980) “Letter to Jourdain”. In: Philosophical and mathematical correspondence, pp78-80. Chicago: Chicago University Press.

Galton CJ, Patterson K, Graham K, Lambon Ralph MA, Williams G, Antoun N, Sahakian BJ, Hodges JR (2001) Differing patterns of temporal atrophy in Alzheimer’s disease and semantic dementia. Neurology 57:216–225.

Goddard E, Carlson TA, Dermody N, Woolgar A (2016) Representational dynamics of object recognition: Feedforward and feedback information flows. NeuroImage 128:385–397.

Gramfort A Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. NeuroImage 86:446–460.

Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37:424-438.

Guggenmos M, Sterzer P, Cichy RM (2018) Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. NeuroImage 173:434-447.

Haggort P (2005) On Broca, brain, and binding: A new framework. Trends Cogn Sci 9:416-423.

Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: Minimum norm estimates. Med Biol Eng Comput 32:35–42.

Hirabayashi T, Tamura K, Takeuchi D, Takeda M, Koyano KW, Miyashita Y (2014) Distinct neuronal interactions in anterior inferotemporal areas of macaque monkeys during retrieval of object association memory. J Neurosci 34:9377-9388.

Hodges JR, Patterson K, Oxbury S, Funnell E (1992) Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. Brain 115:1783–1806.

Hoffman P, McClelland JL, Lambon Ralph MA (2018) Concepts, control and context: A connectionist account of normal and disordered semantic cognition. Psychol Rev 125:293-328.

Ison MJ, Quian Quiroga R, Freid I (2015) Rapid encoding of new memories by individual neurons in the human brain. Neuron 87:220-230.

Jackson RI, Hoffman P, Pobric G, Lambon Ralph MA (2016) The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. J Neurosci 36:1490-1501.

Jang A, Wittig Jr J, Intai S, Zaghloul K (2017) Human cortical neurons in the anterior temporal lobe reinstate spiking activity during verbal memory retrieval. Curr Biol 27:1700–1705.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. Nat Rev Neurosci 18:42-55.

Lyu B, Choi HS, Marslen-Wilson WD, Clarke A, Randall B, Tyler LK (2019) Neural dynamics of semantic composition. Proc Natl Acad Sci USA 116:21318–21327.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Meth 164:177–190.

Matchin W, Hickok G (2020) The cortical organization of syntax. Cereb Cortex 30:1481-1498.

Matchin W, Liao C, Gaston P, Lau E (2019) Same words, different structures: An fMRI investigation of argument relations and the angular gyrus. Neuropsychologia 125:116-128.

Miller JF, Neufang M, Solway A, Brandt A, Trippel M, Mader I, Hefft S, Merkow M, Polyn SM, Jacobs J, Kahana MJ, Schulze-Bonhage A (2013) Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science 342:1111-1114.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage 103:103-118.

Mitchell J, Lapata M (2008) Vector-based models of semantic composition. In: Proc of ACL-HLT 8, pp236-244. Columbus, Ohio: Association for Computational Linguistics.

Nestor PJ, Fryer TD, Hodges JR (2006) Declarative memory impairments in Alzheimer's disease and semantic dementia. NeuroImage 30:1010–1020.

Nieuwenhuis I, Takashima A, Oostenveld R, McNaughton BL, Fernández G, Jensen O (2011) The neocortical network representing associative memory reorganizes with time in a process engaging the anterior temporal lobe. Cereb Cortex 22:2622-2633.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97-113.

Patterson K, Nestor P, Rogers T (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci 8:976–987.

Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, Kastman E, Lindeløv JK (2019) PsychoPy2: Experiments in behavior made easy. Behav Res Methods 51:195–203.

Pennington J, Socher R, Manning C (2014) GloVe: Global vectors for word representation. In: Proc of EMNLP, pp1532–1543. Doha, Qatar: Association for Computational Linguistics.

Pobric G, Jefferies E, Lambon Ralph MA (2007) Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. Proc Natl Acad Sci USA 104:20137-20141.

Price A, Bonner MF, Peelle JE, Grossman M (2015) Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. J Neurosci 35:3276-3284.

Pylkkänen L (2019) The neural basis of combinatory syntax and semantics. Science 366:62–66.

Rey HG, De Falco E, Ison MJ, Valentin A, Alarcon G, Selway R, Richardson MP, Quian Quiroga R (2018) Encoding of long-term associations through neural unitization in the human medial temporal lobe. Nat Commun 9:4372.

Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. Nature 354:152-155.

Schölkopf B, Tsuda K, Vert JP (2004) Kernel methods in computational biology. Cambridge: MIT Press.

Simon DA, Lewis G, Marantz A (2017) Disambiguating form and lexical frequency effects in MEG responses using homonyms. Lang Cognitive Proc 27:275-287.

Teige C, Cornelissen PL, Mollo G, del Jesus Gonzalez Alama TR, McCarty K, Smallwood J, Jefferies E (2019) Dissociations in semantic cognition: Oscillatory evidence for opposing effects

of semantic control and type of semantic relation in anterior and posterior temporal cortex. Cortex 120:308-325.

Teige C, Mollo G, Millman R, Savill N, Smallwood J, Cornelissen PL, Jefferies E (2018) Dynamic semantic cognition: Characterising coherent and controlled conceptual retrieval through time using magnetoencephalography and chronometric transcranial magnetic stimulation. Cortex 103:329-349.

van der Maaten LJP (2014) Accelerating t-SNE using tree-based algorithms. J Mach Learn Res 15:3221-3245.

Wang Y, Collins JA, Koski J, Nugiel T, Metoki A, Olson IR (2017) Dynamic neural architecture for social knowledge retrieval. Proc Natl Acad Sci USA 114:E3305-E3314.

Westerlund M, Kastner I, Al Kaabi M, Pylkkänen L (2015) The LATL as locus of composition: MEG evidence from English and Arabic. Brain Lang 141:124–134.

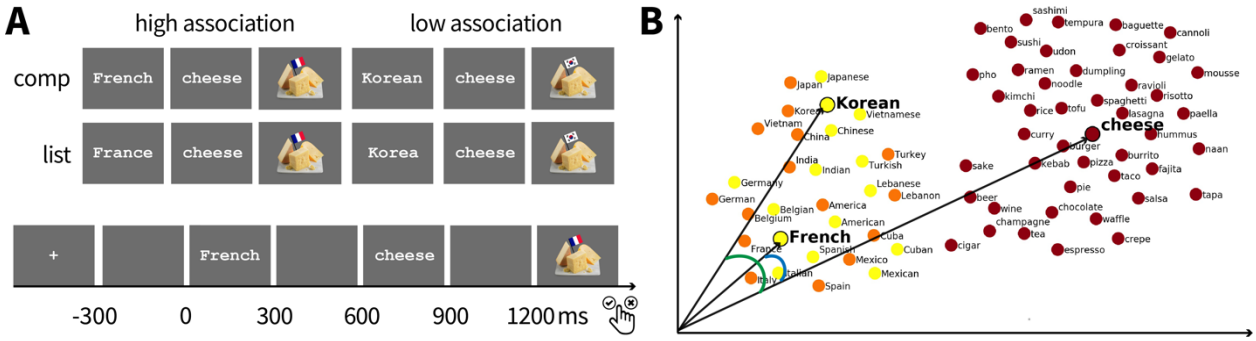


Figure 1. Experimental design. ***A***, Experimental design and trial structure. Our design crossed strength of association (Low vs High) and compositionality (List vs Comp). In each trial, participants indicated whether the target picture matched the preceding words. Half of the target pictures matched and half did not. Activities recorded from 100 ms pre-stimuus onset to 1200 ms post-stimulus onset were analyzed. ***B***, Schematic diagram showing how cosine value between high-dimensional vectors represents semantic similarity. The smaller the angle is between two vectors, the higher the cosine value and semantic similarity. The angle between "French" and "cheese" is smaller than the one between "Korean" and "cheese" because they share more contextual features. The high-dimensional word embeddings were visualized on the 2D scatter using *t*-Distributed Stochastic Neighbor Embedding (t-SNE).

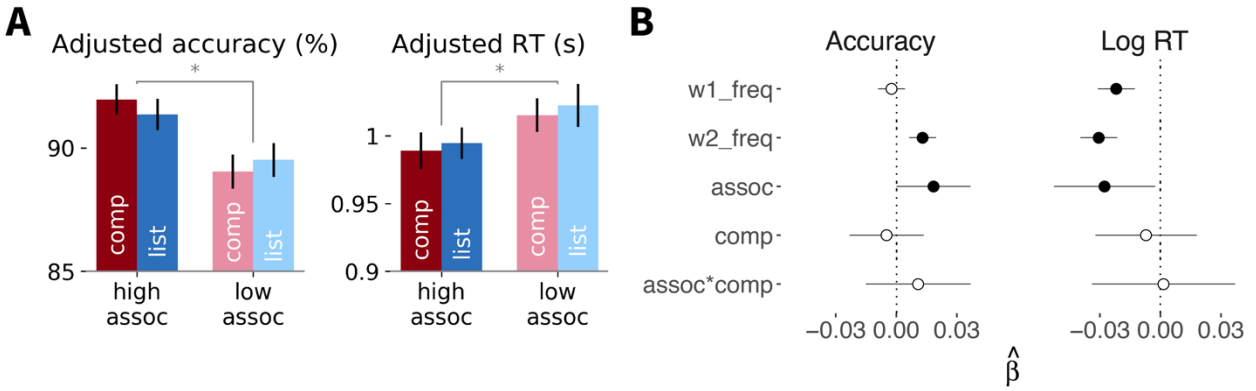


Figure 2. Behavioral results. ***A***, Mean predicted accuracy and reaction time for the four conditions, regressing out the effect of word frequency. Association is a significant predictor for both accuracy (p=0.049) and reaction time (*p*=0.03). Error bars indicate 1 standard error. * *p*<.05**. *B,*** Fitted coefficients for all predictors for accuracy and log reaction time. Error bars show 95% confidence intervals. Filled points indicate statistically significant coefficients at the level of *p*<.05.

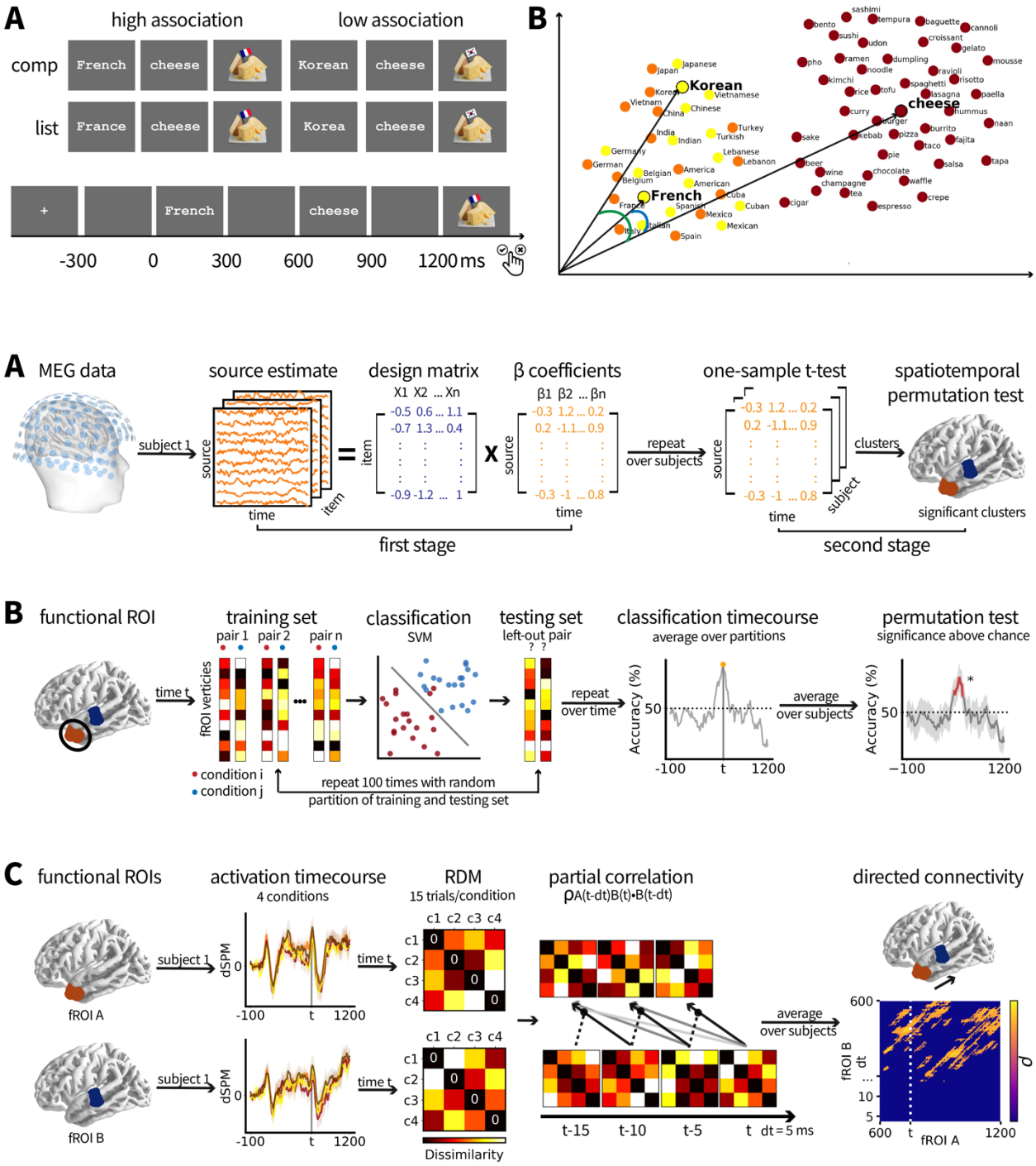


Figure 3. Schematic illustration of the analysis procedure. ***A***, Two-stage multiple regression analyses. At the first stage, an ordinary least squares regression was applied to each participant's single-trial source estimates for each source within a selected region at each timepoint of the analysis window. At the second stage, a one-sample *t*-test was performed on the distribution of β values across subjects for each variable at each source and each timepoint, to test

if their values were significantly different from zero. The significant clusters were determined by a cluster-based spatiotemporal permutation test with 10,000 permutations. ***B***, Multivariate pattern classification analyses. A linear Support Vector Machine (SVM) was trained on the combination of pseudo-trials from 2 conditions and tested on a left-out pair with 100 permutations. The same SVM analysis was applied independently to each timepoint and for each subject. Classification accuracy averaged over subjects at each timepoint minus the chance level of 50% was submitted to a cluster-based one-sample *t*-test with 10,000 permutations. ***C***, Directed connectivity analyses. The representational dissimilarity matrices (RDMs) of the source estimates of the two fROIs at time point *t*, $D(A, t)$ and $D(B, t)$, was calculated as 1 minus the correlation between the conditions. Directed activity from region A to region B was quantified as the partial correlation coefficient between $D(A, t\text{-}dt)$ and $D(B, t)$, partialling out $D(B, t-dt)$, where *dt* is the time interval between the current time point and the previous time point. Significance of coefficients greater than 0 was determined by 10,000 nonparametric cluster-based permutations with a cluster-wise $p < .05$.

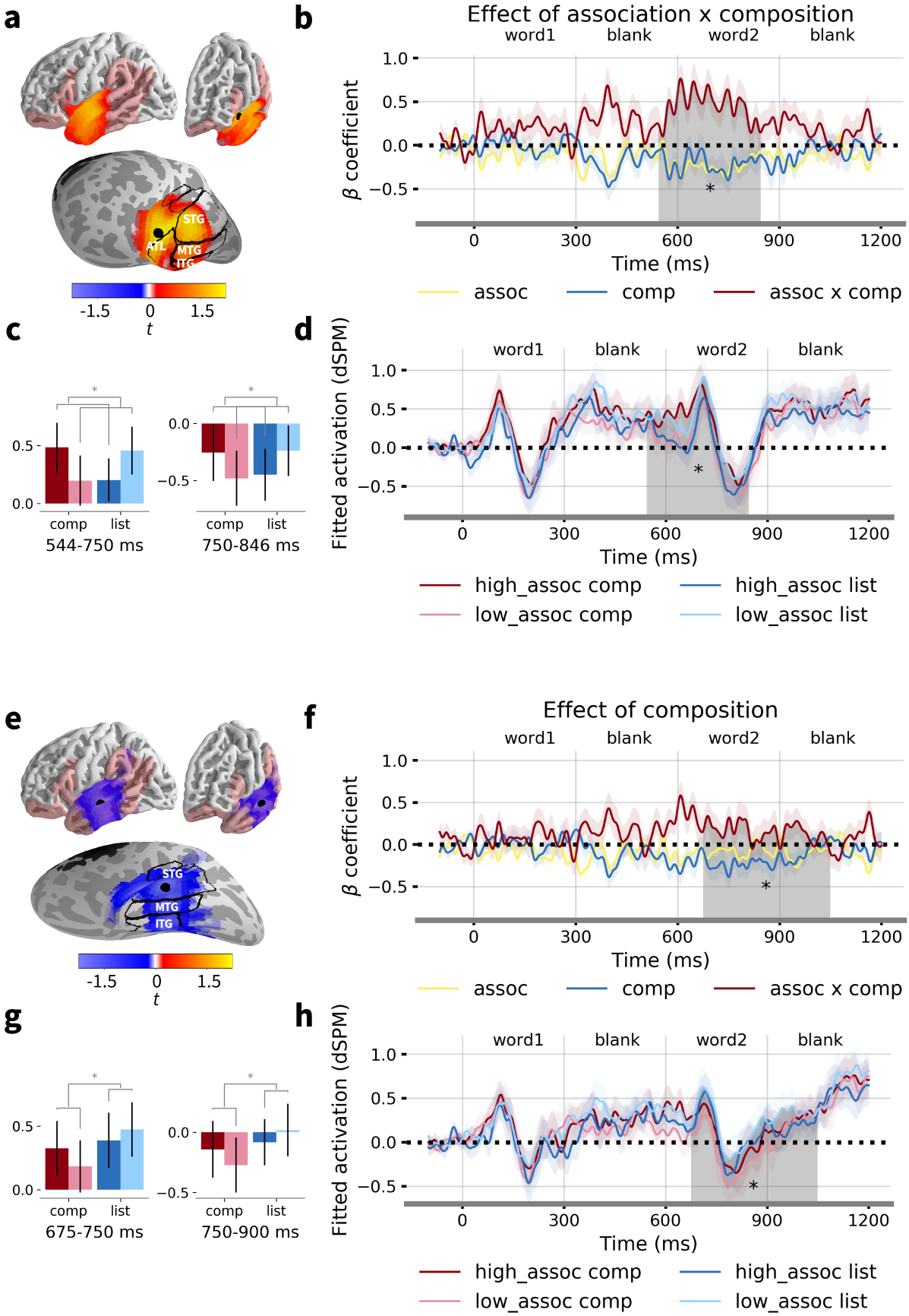

a
STG
ATL
MTG
ITG
-1.5
0
1.5
t
b
Effect of association x composition
word1
blank
word2
blank
β coefficient
1.0
0.5
0.0
−0.5
0
300
600
900
1200
Time (ms)
assoc
comp
assoc x comp
c
0.5
0.0
comp
list
544-750 ms
0.0
−0.5
comp
list
750-846 ms
d
Fitted activation (dSPM)
word1
blank
word2
blank
1.0
0.5
0.0
−0.5
0
300
600
900
1200
Time (ms)
high_assoc comp
high_assoc list
low_assoc comp
low_assoc list
e
STG
MTG
ITG
-1.5
0
1.5
t
f
Effect of composition
word1
blank
word2
blank
β coefficient
1.0
0.5
0.0
−0.5
0
300
600
900
1200
Time (ms)
assoc
comp
assoc x comp
g
0.5
0.0
comp
list
675-750 ms
0.0
−0.5
comp
list
750-900 ms
h
Fitted activation (dSPM)
word1
blank
word2
blank
1.0
0.5
0.0
−0.5
0
300
600
900
1200
Time (ms)
high_assoc comp
high_assoc list
low_assoc comp
low_assoc list

Figure 4. Multiple regression results. ***A***, Location of the significant cluster sensitive to the interaction effect between association and composition derived from the spatiotemporal cluster test. Light red regions mark the language network within which the analysis was conducted. Black dot indicates location of peak activity at the latency of 608 ms. Color bar indicates *t*-statistics. ***B***, Timecourses of β coefficients for association, composition and their interaction term averaged over the cluster. ***C***, Fitted responses for each condition averaged over the cluster and the time window of 544-750 ms for positive polarity and of 750-846 ms for negative polarity. ***D***, Timecourses of fitted response for each condition averaged over the cluster. Word frequency effects were regressed out of the responses. Shaded region indicates the significant time window from 544-846 ms (p=0.019). ***E***, Location of the significant cluster sensitive to the composition effect derived from the spatiotemporal cluster test. Black dot indicates location of peak activity at the latency of 860 ms. ***F***, Timecourses of β coefficients for association, composition and their interaction term averaged over the cluster. ***G***, Fitted responses for each condition averaged over the cluster and the time window of 675-750 ms for positive polarity and of 750-900 ms for negative polarity. ***H***, Timecourses of fitted response for each condition averaged over the cluster. Word frequency effects were regressed out of the responses. Shaded region indicates the significant time window from 675-1050 ms (p=0.032). * $p < .05$. STG: superior temporal gyrus; MTG: middle temporal gyrus; ITG: inferior temporal gyrus; ATL: anterior temporal lobe.

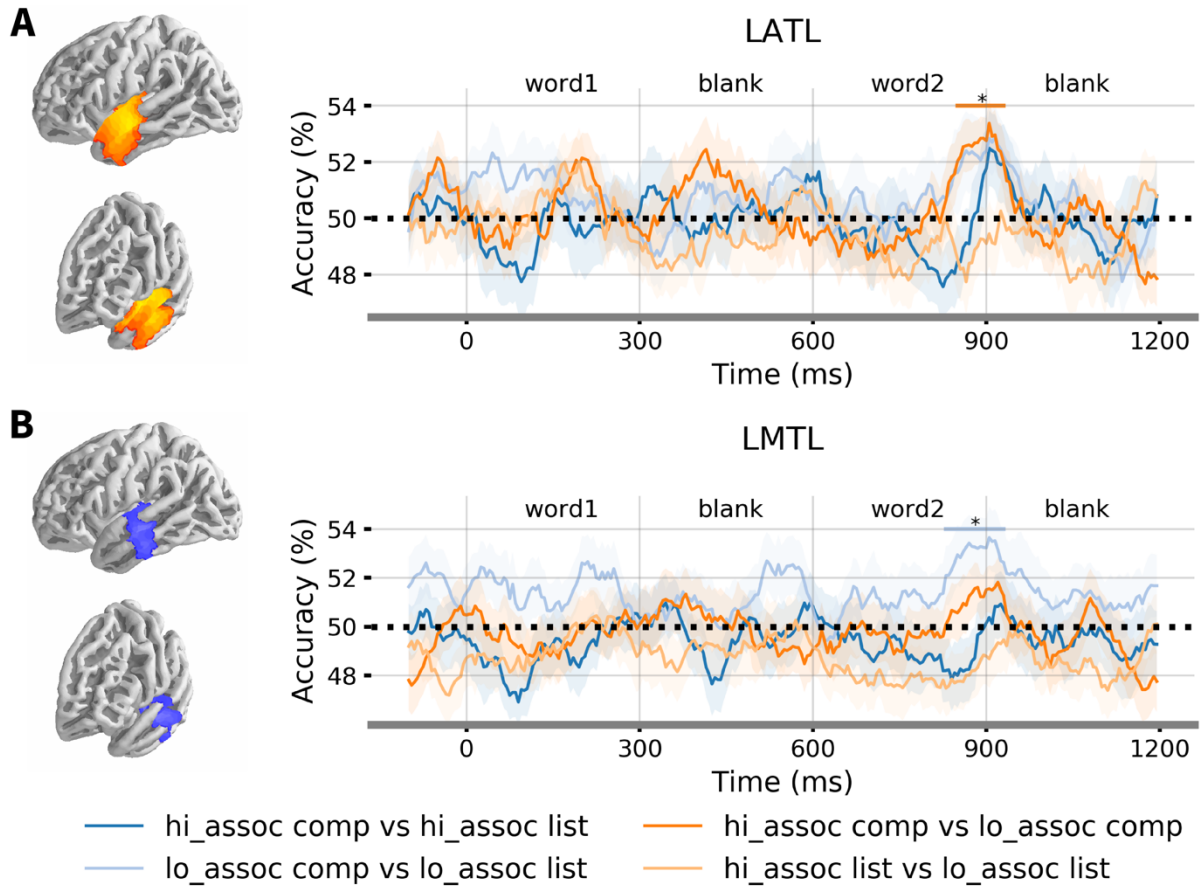


Figure 5. Multivariate pattern classification results. ***A***, Color-coded curves report classification timecourses for the 4 linear SVM classifiers within the functional ROI (thresholded at $|t| > 1$) for the interaction effect between association and composition identified by the regression analyses. Horizontal lines above the curves mark a statistically significant time window from 850-935 ms for the high-association composition vs low-association composition classifier (p=0.019). ***B***, Classification timecourses for the 4 linear SVM classifiers within the functional ROI for the composition effect identified by the regression analysis. Horizontal line above the curves marks a significant time window from 830-935 ms for the low-association composition vs low-association list classifier (p=0.035). Significance over the chance level of 50% was assessed by a cluster-based permutation t-test within the analysis window of 600-1200 ms. * $p < .05$.

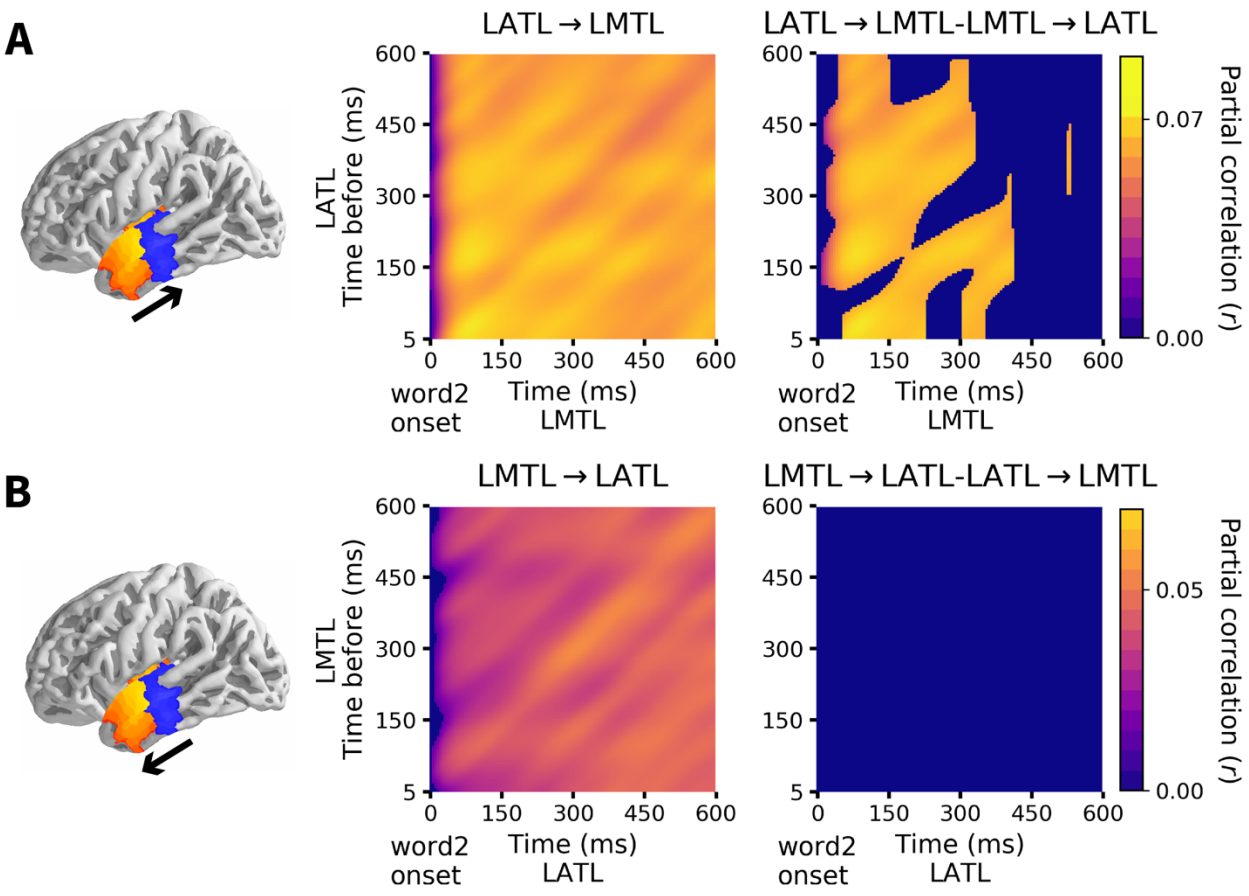


Figure 6. Directed connectivity between the LATL and the LMTL. ***A***, Significant partial correlation coefficients matrix from LATL to LMTL. Contrast between the two directed connectivity measures show significant correlations from LATL to LMTL from 0 to 300 ms after the onset of the second word, with a delay from 5-600 ms. ***B***, Significant partial correlation coefficients matrix from LMTL to LATL. Contrast between the two directed connectivity measures show no significant correlations from LMTL to LATL.

Supplemental Information

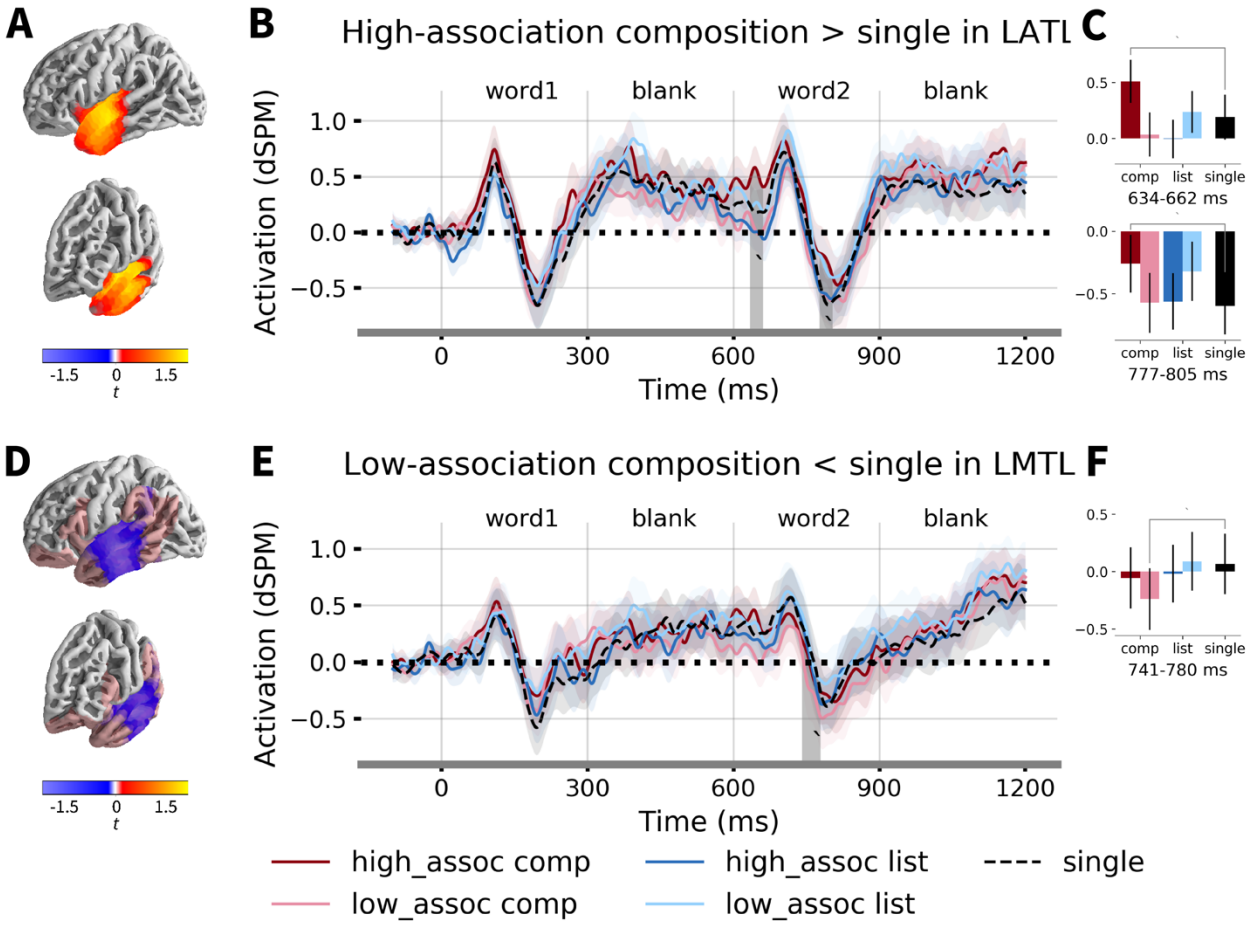
Experiment 1

high association			low association		
adj+noun	noun+noun	cos	adj+noun	noun+noun	cos
American burger	America burger	0.301	American crepes	America crepes	0.088
American pie	America pie	0.344	American dumplings	America dumplings	0.117
Belgian chocolate	Belgium chocolate	0.473	American kebab	America kebab	0.109
Belgian waffle	Belgium waffle	0.455	American paella	America paella	0.074
Chinese dumplings	China dumplings	0.406	American prawn	America prawn	0.061
Chinese noodles	China noodles	0.436	American risotto	America risotto	0.056
Chinese rice	China rice	0.423	American tapas	America tapas	0.082
Chinese tea	China tea	0.413	American tofu	America tofu	0.109
Chinese tofu	China tofu	0.347	American udon	America udon	0.068
Cuban cigar	Cuba cigar	0.547	Chinese cannoli	China cannoli	0.033
French baguette	France baguette	0.39	Chinese espresso	China espresso	0.12
French champagne	France champagne	0.4	English bibimbap	England bibimbap	-0.007
French cheese	France cheese	0.458	English fajitas	England fajitas	-0.017
French crepes	France crepes	0.386	English guacamole	England guacamole	0.04
French croissant	France croissant	0.419	English hummus	England hummus	0.058
French mousse	France mousse	0.305	English kimchi	England kimchi	0.088
French toast	France toast	0.381	English pho	England pho	0.112
French wine	France wine	0.38	English ravioli	England ravioli	0.063
German beer	Germany beer	0.314	English taco	England taco	0.155
Greek salad	Greece salad	0.396	English tequila	England tequila	0.113
Greek yogurt	Greece yogurt	0.391	English tortilla	England tortilla	0.083
Indian curry	India curry	0.429	German salsa	Germany salsa	0.179
Indian naan	India naan	0.305	German sashimi	Germany sashimi	0.043
Irish whiskey	Ireland whiskey	0.474	German yogurt	Germany yogurt	0.124
Italian cannoli	Italy cannoli	0.319	Greek cigar	Greece cigar	0.101
Italian cappuccino	Italy cappuccino	0.343	Indian baguette	India baguette	0.067
Italian espresso	Italy espresso	0.349	Indian cappuccino	India cappuccino	0.054
Italian gelato	Italy gelato	0.413	Indian croissant	India croissant	0.032
Italian lasagna	Italy lasagna	0.427	Indian gelato	India gelato	0.025
Italian latte	Italy latte	0.316	Indian lasagna	India lasagna	0.113
Italian pasta	Italy pasta	0.523	Indian mousse	India mousse	0.03
Italian pizza	Italy pizza	0.559	Indian vodka	India vodka	0.139
Italian ravioli	Italy ravioli	0.382	Indian waffle	India waffle	0.088
Italian risotto	Italy risotto	0.357	Irish bento	Ireland bento	0.038
Italian spaghetti	Italy spaghetti	0.453	Irish ramen	Ireland ramen	0.122
Japanese bento	Japan bento	0.318	Japanese naan	Japan naan	0.079
Japanese ramen	Japan ramen	0.38	Korean champagne	Korea champagne	0.129
Japanese sake	Japan sake	0.293	Korean cheese	Korea cheese	0.163
Japanese sashimi	Japan sashimi	0.338	Korean latte	Korea latte	0.131
Japanese sushi	Japan sushi	0.505	Korean whiskey	Korea whiskey	0.144
Japanese tempura	Japan tempura	0.3	Lebanese beer	Lebanon beer	0.122
Japanese udon	Japan udon	0.319	Lebanese chocolate	Lebanon chocolate	0.143
Korean bibimbap	Korea bibimbap	0.387	Lebanese tea	Lebanon tea	0.105
Korean kimchi	Korea kimchi	0.46	Russian burger	Russia burger	0.133
Lebanese hummus	Lebanon hummus	0.431	Russian burrito	Russia burrito	0.063
Mexican burrito	Mexico burrito	0.515	Russian curry	Russia curry	0.172
Mexican fajitas	Mexico fajitas	0.434	Russian laksa	Russia laksa	0.012
Mexican guacamole	Mexico guacamole	0.443	Russian pasta	Russia pasta	0.139
Mexican salsa	Mexico salsa	0.506	Russian spaghetti	Russia spaghetti	0.135
Mexican taco	Mexico taco	0.626	Scottish pizza	Scotland pizza	0.192
Mexican tequila	Mexico tequila	0.464	Scottish sushi	Scotland sushi	0.157
Russian vodka	Russia vodka	0.32	Singaporean pie	Singapore pie	0.081
Scottish haggis	Scotland haggis	0.479	Singaporean rice	Singapore rice	0.146
Singaporean laksa	Singapore laksa	0.419	Singaporean salad	Singapore salad	0.136
Spanish paella	Spain paella	0.353	Singaporean toast	Singapore toast	0.067
Spanish tapas	Spain tapas	0.354	Singaporean wine	Singapore wine	0.018
Spanish tortilla	Spain tortilla	0.312	Spanish haggis	Spain haggis	0.106
Thai prawn	Thailand prawn	0.4	Spanish noodles	Spain noodles	0.139
Turkish kebab	Turkey kebab	0.419	Spanish tempura	Spain tempura	0.082
Vietnamese pho	Vietnam pho	0.616	Turkish sake	Turkey sake	0.102

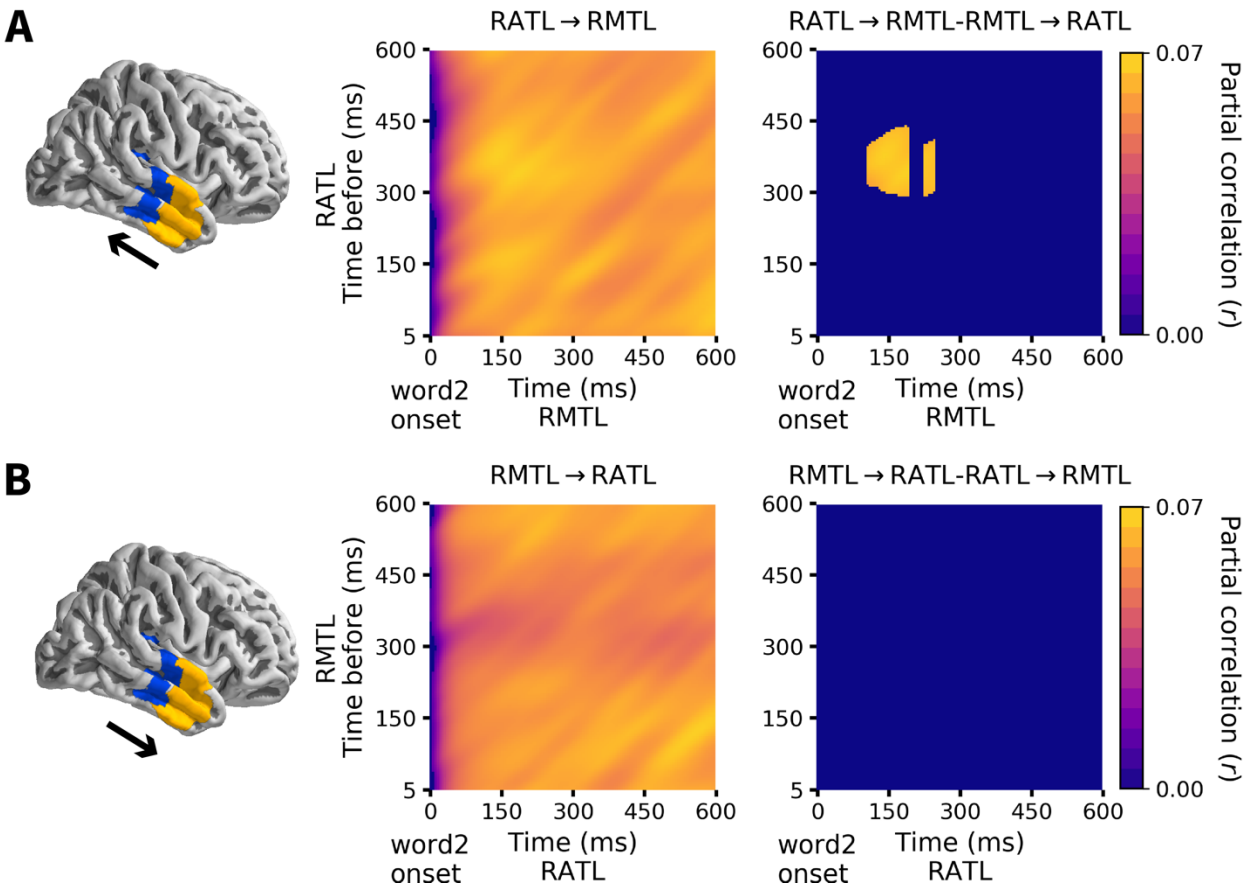
Experiment 2

high association			low association		
adj+noun	noun+noun	cos	adj+noun	noun+noun	cos
American burger	America burger	0.301	American crepe	America crepe	0.115
American pie	America pie	0.344	American kebab	America kebab	0.109
Belgian chocolate	Belgium chocolate	0.473	American ravioli	America ravioli	0.075
Belgian waffle	Belgium waffle	0.455	American risotto	America risotto	0.056
Chinese dumpling	China dumpling	0.388	American tapa	America tapa	0.019
Chinese noodle	China noodle	0.470	Belgian pizza	Belgium pizza	0.246
Chinese rice	China rice	0.423	Belgian spaghetti	Belgium spaghetti	0.153
Chinese tea	China tea	0.413	Chinese cannoli	China cannoli	0.033
Chinese tofu	China tofu	0.347	Chinese espresso	China espresso	0.120
Cuban cigar	Cuba cigar	0.547	Chinese hummus	China hummus	0.139
French baguette	France baguette	0.390	Cuban pho	Cuba pho	0.236
French champagne	France champagne	0.400	Cuban tofu	Cuba tofu	0.158
French cheese	France cheese	0.458	Cuban udon	Cuba udon	0.104
French crepe	France crepe	0.333	German burrito	Germany burrito	0.089
French croissant	France croissant	0.419	German curry	Germany curry	0.190
French mousse	France mousse	0.305	German kimchi	Germany kimchi	0.119
French wine	France wine	0.380	German rice	Germany rice	0.139
German beer	Germany beer	0.314	German salsa	Germany salsa	0.179
Indian curry	India curry	0.429	German sashimi	Germany sashimi	0.043
Indian naan	India naan	0.305	German sushi	Germany sushi	0.177
Italian cannoli	Italy cannoli	0.319	Indian baguette	India baguette	0.067
Italian espresso	Italy espresso	0.349	Indian croissant	India croissant	0.032
Italian gelato	Italy gelato	0.413	Indian gelato	India gelato	0.025
Italian lasagna	Italy lasagna	0.427	Indian lasagna	India lasagna	0.113
Italian pizza	Italy pizza	0.559	Indian mousse	India mousse	0.030
Italian ravioli	Italy ravioli	0.382	Indian waffle	India waffle	0.088
Italian risotto	Italy risotto	0.357	Japanese fajita	Japan fajita	0.062
Italian spaghetti	Italy spaghetti	0.453	Japanese naan	Japan naan	0.079
Japanese bento	Japan bento	0.318	Korean cheese	Korea cheese	0.163
Japanese ramen	Japan ramen	0.380	Korean paella	Korea paella	0.139
Japanese sake	Japan sake	0.293	Korean wine	Korea wine	0.137
Japanese sashimi	Japan sashimi	0.338	Lebanese beer	Lebanon beer	0.122
Japanese sushi	Japan sushi	0.505	Lebanese chocolate	Lebanon chocolate	0.143
Japanese tempura	Japan tempura	0.300	Lebanese pie	Lebanon pie	0.257
Japanese udon	Japan udon	0.319	Lebanese tea	Lebanon tea	0.105
Korean kimchi	Korea kimchi	0.460	Mexican sake	Mexico sake	0.149
Lebanese hummus	Lebanon hummus	0.431	Spanish dumpling	Spain dumpling	0.103
Mexican burrito	Mexico burrito	0.515	Spanish noodle	Spain noodle	0.157
Mexican fajita	Mexico fajita	0.408	Spanish ramen	Spain ramen	0.155
Mexican salsa	Mexico salsa	0.506	Spanish tempura	Spain tempura	0.082
Mexican taco	Mexico taco	0.626	Turkish bento	Turkey bento	0.054
Spanish paella	Spain paella	0.353	Turkish burger	Turkey burger	0.177
Spanish tapa	Spain tapa	0.223	Turkish taco	Turkey taco	0.189
Turkish kebab	Turkey kebab	0.419	Vietnamese champagne	Vietnam champagne	0.170
Vietnamese pho	Vietnam pho	0.616	Vietnamese cigar	Vietnam cigar	0.168

Supplementary Table 1. Stimuli lists for experiment 1 and experiment 2.



Supplementary Figure 1. Pair-wise *t*-test results between each of the two-word conditions and the single word condition within the LATL and the LMTL fROIs. ***A***, The LATL fROI. ***B***, Time course of responses of for each condition averaged over the cluster. Shaded region indicates the marginally significant time windows of 634-662 ms and 777-805 ms for high-association composition > single (p=0.073 and p=0.087, respectively). ***C*,** Fitted responses for each condition averaged over the cluster and the time windows. ***D***, The LMTL fROI. ***E***, Time course of responses for each condition averaged over the cluster. Shaded region indicates the marginally significant time windows of 741-780 ms for low-association composition < single (p=0.072). ***F*,** Fitted responses for each condition averaged over the cluster and the time window. ` p<.1.



Supplementary Figure 2. Directed connectivity between the RATL and the RMTL. ***A***, Significant partial correlation coefficients matrix from RATL to RMTL. Contrast between the two directed connectivity measures show significant correlations from RATL to RMTL from around 150 to 250 ms after the onset of the second word, with a delay from 300-450 ms. ***B***, Significant partial correlation coefficients matrix from RMTL to RATL. Contrast between the two directed connectivity measures show no significant correlations from RMTL to RATL.