

# Decoding the Cortical Dynamics of Sound-Meaning Mapping

Ece Kocagoncu, Alex Clarke, Barry J. Devereux, and Lorraine K. Tyler

Centre for Speech, Language and the Brain, Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom

Comprehending speech involves the rapid and optimally efficient mapping from sound to meaning. Influential cognitive models of spoken word recognition (Marslen-Wilson and Welsh, 1978) propose that the onset of a spoken word initiates a continuous process of activation of the lexical and semantic properties of the word candidates matching the speech input and competition between them, which continues until the point at which the word is differentiated from all other cohort candidates (the uniqueness point, UP). At this point, the word is recognized uniquely and only the target word's semantics are active. Although it is well established that spoken word recognition engages the superior (Rauschecker and Scott, 2009), middle, and inferior (Hickok and Poeppel, 2007) temporal cortices, little is known about the real-time brain activity that underpins the computations and representations that evolve over time during the transformation from speech to meaning. Here, we test for the first time the spatiotemporal dynamics of these processes by collecting MEG data while human participants listened to spoken words. By constructing quantitative models of competition and access to meaning in combination with spatiotemporal searchlight representational similarity analysis (Kriegeskorte et al., 2006) in source space, we were able to test where and when these models produced significant effects. We found early transient effects ~400 ms before the UP of lexical competition in left supramarginal gyrus, left superior temporal gyrus, left middle temporal gyrus (MTG), and left inferior frontal gyrus (IFG) and of semantic competition in MTG, left angular gyrus, and IFG. After the UP, there were no competitive effects, only target-specific semantic effects in angular gyrus and MTG.

**Key words:** cohort; competition; MEG; MVPA; searchlight; semantics

## Significance Statement

Understanding spoken words involves complex processes that transform the auditory input into a meaningful interpretation. This effortless transition occurs on millisecond timescales, with remarkable speed and accuracy and without any awareness of the complex computations involved. Here, we reveal the real-time neural dynamics of these processes by collecting data about listeners' brain activity as they hear spoken words. Using novel statistical models of different aspects of the recognition process, we can locate directly which parts of the brain are accessing the stored form and meaning of words and how the competition between different word candidates is resolved neurally in real time. This gives us a uniquely differentiated picture of the neural substrate for the first 500 ms of word recognition.

## Introduction

Understanding speech involves the rapid transformation from low-level acoustic-phonetic analysis to complex meaning representations. This rapid speech-meaning mapping, which is typically within 200 ms of word onset, is the result of a seemingly

effortless set of computations and is vital for rapid and efficient communication. Although current neurobiological models situate the speech-mapping process in the superior (Rauschecker and Scott, 2009), middle, and inferior (Hickok and Poeppel, 2007) temporal cortices, little is known about the neural dynamics of the computations and representations that evolve over time during the transformation from speech to meaning. Here, we used MRI-constrained MEG in source space combined with quantitative models of cognitive computations and multivariate pattern analysis (MVPA) methods to determine the spatiotemporal dynamics of the computations and representations that evolve over time during the transformation from speech to meaning.

Some influential cognitive models claim that, as speech is heard (e.g., the initial phonemes /rd/ in *robin*), the lexical and semantic properties of word candidates (e.g., *robin*, *rock*, *rod*) that match the incoming speech are rapidly and partially activated, forming the “word-initial cohort” (Marslen-Wilson and Welsh,

Received Sept. 9, 2016; revised Dec. 16, 2016; accepted Dec. 21, 2016.

Author contributions: E.K., B.J.D., and L.K.T. designed research; E.K. and A.C. performed research; E.K., A.C., and B.J.D. contributed unpublished reagents/analytic tools; E.K. analyzed data; E.K. and L.K.T. wrote the paper.

This work was supported by a European Research Council Advanced Investigator grant under the European Community's Horizon 2020 Research and Innovation Programme (2014–2020 ERC Grant 669820 to L.K.T.). We thank Elisa Carrus, Billi Randall, and Elisabeth Fonteneau for invaluable help in data collection and William Marslen-Wilson for constructive discussions on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Lorraine K. Tyler, Centre for Speech, Language and the Brain, Department of Psychology, University of Cambridge Downing Street, Cambridge CB2 3EB, UK. E-mail: l.k.tyler@csl.psychol.cam.ac.uk.

DOI:10.1523/JNEUROSCI.2858-16.2016

Copyright © 2017 the authors 0270-6474/17/371312-08\$15.00/0

1978; McClelland and Elman, 1986; Gaskell and Marslen-Wilson, 2002; Zhuang et al., 2014). The parallel activation of these lexical representations creates transient lexical and semantic competition between candidates that diminishes over time as the pool of candidates narrows to a single target item (the uniqueness point, UP), boosting its specific semantics. This set of processes constitutes an optimally efficient processing system in which a spoken word is identified as soon as it can be differentiated reliably from its neighbors (Marslen-Wilson, 1984, 1987; Zhuang et al., 2014).

To test for the neural spatiotemporal properties of these computations from the initial stages of cohort competition through to the activation of target word semantics and the crucial role of the UP in marking a reduction in competitive processes and increasingly selective activation of the target word's semantics, we collected MEG data while participants listened to 296 familiar words interspersed with a small number of nonwords and pressed a response key when they heard a nonword, thus reducing task effects while maintaining the participant's attention.

We determined empirically the UP of each word using a gating task (Grosjean, 1980) in which listeners heard increasingly larger fragments of a word and guessed its identity. Subjects' gating responses were also used to develop quantitative models of phonologically driven lexical and semantic competition before each word's UP. Lexical competition was defined as the change in lexical uncertainty (Shannon entropy) as the speech unfolds over time and cohort membership changes up to the UP. Semantic competition was derived from the semantic similarity of each word's competitors obtained from the gating data and computed using a large corpus-based distributional memory (DM) database (Baroni and Lenci, 2010). We derived a quantifiable measure of target-specific semantics from a large, semantic feature-based corpus (McRae et al., 2005) using the relationship between two feature-specific indices, feature distinctiveness and correlational strength, which were shown in previous studies to be a key property of lexical semantics (Randall et al., 2004; Taylor et al., 2012; Tyler et al., 2013; Devereux et al., 2016).

The MEG analyses used spatiotemporal searchlight representational similarity analysis (ssRSA; Su et al., 2012) in MEG source space. ssRSA compares the similarity structure observed in brain activity time courses with the similarity structure of our theoretically relevant cognitive models. RSA can reveal distinct representational geometries in different brain areas even when other MVPA methods fail (Connolly et al., 2012). Using ssRSA, we constructed model representational dissimilarity matrices (RDMs) of our cognitive variables: lexical and semantic competition between activated cohort members and target-specific semantic information. We tested these over time and against activity patterns in MEG source space (captured as brain data RDMs) focusing specifically before and after the UP. We predicted early pre-UP lexical and semantic competition effects and later post-UP target-specific semantic effects.

## Materials and Methods

**Participants.** Eleven healthy participants (mean age: 24.4 years, range 19–35; 7 females, 4 males) volunteered to participate in the study. All were right-handed, native British English speakers with normal hearing. The experiment was approved by the Cambridge Psychology Research Ethics Committee.

**Procedure.** Participants were seated in a magnetically shielded room (IMEDCO) and positioned under the MEG scanner. Auditory stimuli were delivered binaurally through MEG-compatible ER3A insert ear-

phones (Etymotic Research). The delay in sound delivery due to the length of earphone tubes and stimulus delivery computer's sound card was  $32 \pm 2$  ms on average. Instructions were presented visually on the screen positioned 1 m in front of the participant's visual field.

Before the E/MEG recording, participants performed an automated hearing test in the MEG scanner, where they were presented with pure tones at a frequency of 1000 Hz to either ear and pressed a response button when they heard a tone. The loudness of the tones was gradually attenuated until the participant was unable to hear them. By this means, we determined hearing thresholds for each ear. Participants' hearing thresholds (right ear  $M = 66.71$ ,  $SD = 9.95$ ; left  $M = 62.5$ ;  $SD = 11.21$ ) were all within the normal range (45–75 dB).

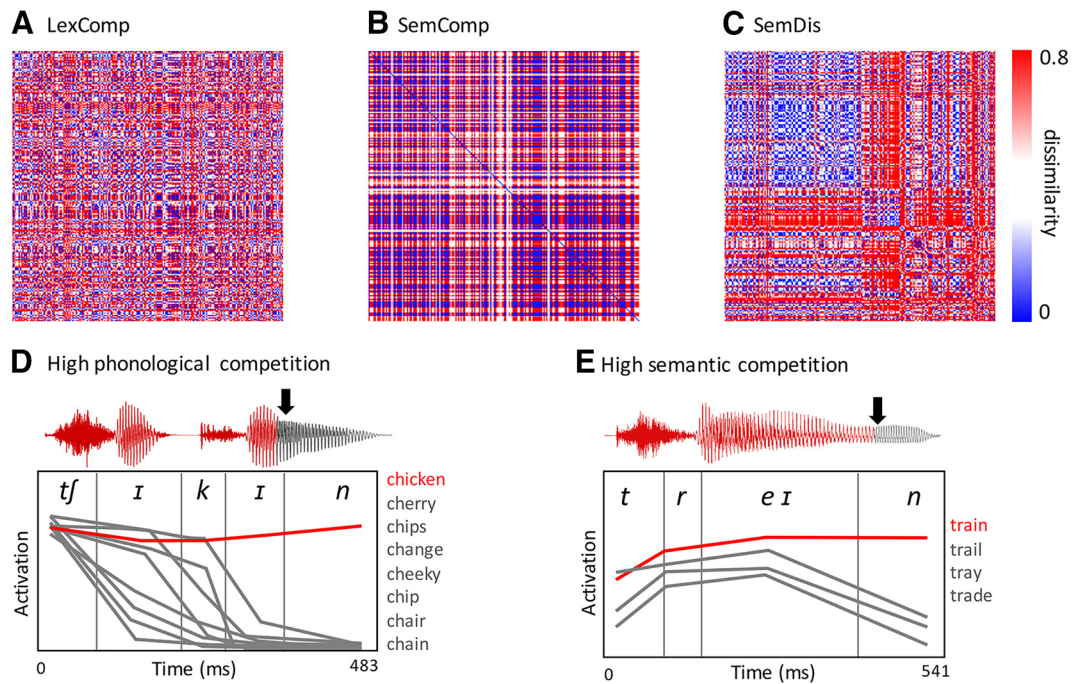
To ensure that participants were listening attentively to the stimuli, a simple nonword detection task was performed on 10% of the trials. Participants were instructed to listen to the speech stimuli and press a response key whenever they heard a nonword. Participants were instructed to focus their eyes on a fixation cross that appeared for 650 ms before the onset of the auditory stimulus. The interstimulus interval was randomized between 1500 and 2500 ms. Every interval was followed by a blink break that lasted for 1500 ms. The start of the blink break was indicated by an image of an eye that appeared in the middle of the screen. Using blink breaks and fixation cross, we aimed to minimize the contamination of the signal by eye-movement-related artifacts. Participants were also asked to refrain from movement during the entire block of recording. E-Prime 2 (Psychology Software Tools) was used to present the stimuli and record participants' behavioral responses to the nonword detection task.

**Stimuli.** The stimuli consisted of 218 spoken names of concrete objects (e.g., *alligator, hammer, cabbage*) and 30 phonotactically legal nonwords (e.g., *rayber, chickle, tomula*). Nonwords were matched to real words by their duration, number of syllables, and phonemes. The nonword trials were excluded from the imaging analysis. All of the words were highly familiar ( $M = 477$ ,  $SD = 73$ ; MRC Psycholinguistic Database), frequent ( $M = 19.21$ ,  $SD = 39.01$ ; CELEX; Baayen et al., 1993), semantically rich (number of features,  $M = 13$ ,  $SD = 3.34$ ) concepts with a mean duration of 602 ms ( $SD = 122$  ms).

The stimuli were presented in a pseudorandomized order such that consecutive items were from different domains of knowledge (living or nonliving concepts) and started with different phonemes. They were presented in two blocks, each containing 109 words and 15 nonwords. The block order was also randomized for each participant.

**Gating task.** We used a behavioral gating task (Grosjean, 1980; Tyler and Wessels, 1985) to determine the UP of each spoken word and to obtain word candidates to serve as inputs to the generation of models of lexical and semantic competition. In a self-paced procedure, 45 participants, who did not take part in the MEG study, listened to incremental segments (i.e., gates; e.g., *al. . . alli. . . allig. . .*) of an initial set of 372 spoken words and were asked to type in their best guess of the word and rate their confidence in their answer on a scale of 1–7 (where 7 = very confident and 1 = not confident at all). We initially used CELEX (Baayen et al., 1993) to determine an approximate UP to guide the onset of the gating segments. We gated before and after the UP in five increments of 25 ms each. The gating UPs were defined as the gate where 80% of the participants correctly identified the word with an average confidence rating of at least 80% (Warren and Marslen-Wilson, 1987, 1988). Using these criteria, we determined the gating UPs of the 218 spoken words that were presented in the MEG experiment (time from word onset;  $M = 408$  ms,  $SD = 81$  ms) and obtained a list of the word candidates produced by subjects at each gate. The gating UPs were on average 69 ms later than the CELEX UPs.

**Cognitive models.** We modeled the key cognitive processes of lexical competition, semantic competition, and the ease of semantic feature integration into a coherent concept. The model of lexical competition, the LexComp model (Fig. 1A), was defined as the change in cohort entropy from the first gate until the UP. Here, entropy represents the lexical uncertainty that results from changing lexical representations as speech unfolds and is calculated with Shannon's entropy formula (Shannon and Weaver, 1949; Shannon, 2001):



**Figure 1.** Model RDMs tested in the current analysis where rows and columns of the matrices represent trials (i.e., single spoken words) and values indicate pairwise dissimilarity values across trials. **A**, LexComp model. **B**, SemComp model. **C**, SemDis model. **D** and **E** represent cohort activations of two example stimuli taken from the gating results: *chicken* and *train*. Cohort profiles indicate the change in activation of every candidate word over time. **D**, At word onset, all members in *chicken*'s cohort are partially activated, resulting in high lexical competition. As more of the speech input is heard, the activation of competitor words decays and the target word's activation increases. **E**, *Train*'s cohort profile shows high semantic competition due to low semantic overlap between the cohort candidates. Arrows mark the UPs.

$$H = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

where  $H$  refers to entropy,  $P(x_i)$  refers to the summed confidence score for a cohort competitor  $i$  (e.g., the target word *train* has *trade*, *trail*, and *tray* as cohort competitors) across participants divided by the total sum of confidence scores for all the competitors reported at a particular gate and  $n$  refers to the total number of competitors at that gate. Here, differences in entropy were inversely related to the level of lexical competition.

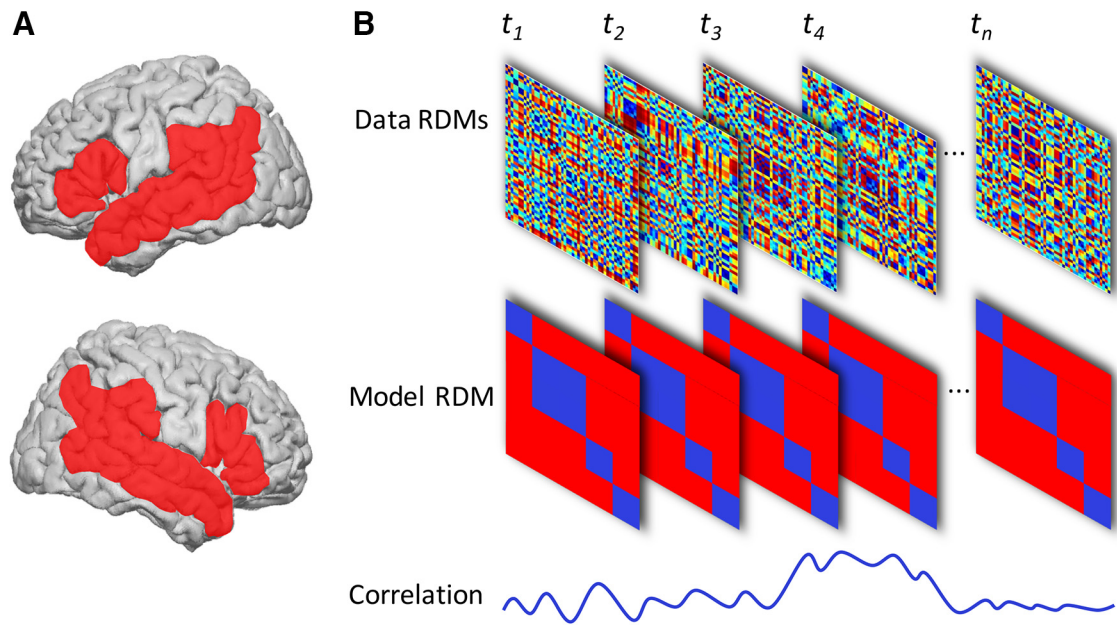
Semantic competition (the SemComp model) incorporated semantic information about all competitor words included in the target word's cohort (Fig. 1B) as determined by the gating data. The semantic similarity of the set of each word's competitors was computed using a corpus-based DM database (Baroni and Lenci, 2010). The DM database represents the semantics of words as vectors over 5000 semantic dimensions, where the entries of the semantic dimensions are derived from word cooccurrence data dimensionally reduced by random indexing (Kanerva et al., 2000). Semantic competition was defined as the average cosine similarity (i.e., semantic overlap) between pairs of vectors of competitors at the first gate. A smaller overlap between a pair of vectors was proposed to create an average pattern of semantic activation from both words that has a small resemblance to any single conceptual representation and therefore generating higher semantic competition (Gaskell and Marslen-Wilson, 1997).

Finally, the SemDis model (Fig. 1C) was designed to tap into the ease with which the semantic features of a concept are integrated to form a unique meaning representation of each target word. This measure was computed using the relationship between two feature-specific indices: feature distinctiveness and correlational strength (Randall et al., 2004; Taylor et al., 2012; Tyler et al., 2013; Devereux et al., 2014; i.e., feature cooccurrence). The set of features for each word were obtained from a database of anglicized feature norms (McRae et al., 2005; Taylor et al., 2012). Feature distinctiveness was calculated as:  $1/(\text{number of concepts in which a feature occurred})$  (e.g., *has stripes*, *has a hump*). Correlational strength measured the degree to which a feature cooccurs with other features. For each concept, we can plot its constituent features in a scatterplot with feature correlational strength on the  $x$ -axis and feature dis-

tinctiveness on the  $y$ -axis (Taylor et al., 2012). The SemDis measure was defined as the unstandardized slope of the regression line fitted to this plot. The *SemDis* measure captured the relative contribution of a concept's feature cooccurrence to feature distinctiveness, so it was sensitive to the ease of feature integration of concepts (Taylor et al., 2012).

**MEG and MRI acquisition.** Continuous MEG data were recorded using the whole-head 306-channel Vector-view system (Electa Neuromag). The channel system consisted of 102 pairs of planar gradiometers and 102 magnetometers. To monitor and record blinks and cardiac activity, EOG and ECG electrodes were used. To record subjects' head shape and movements, five head position indicator (HPI) coils were attached onto the subjects' head. HPI coils recorded the head position every 200 ms. For coregistration of the subject's head to the MEG sensors, the three fiducial points (nasion, left, and right pre-auricular) and additional points across the scalp were digitized. MEG signals were recorded continuously at a 1000 Hz sampling rate with a high-pass filter of 0.03 Hz. To facilitate source localization, T1-weighted MP-RAGE scans with 1 mm isotropic resolution were acquired for each subject using Siemens 3 T Tim Trio. Both the MEG and MRI systems were located at the Medical Research Council Cognition and Brain Sciences Unit in Cambridge, United Kingdom.

**MEG preprocessing and source localization.** The raw data were processed using MaxFilter 2.2 (Elekta Oy) in three steps. In the first stage, bad channels were detected and reconstructed using interpolation. In the second stage, signal space separation was applied to the data every 4 s to separate the signals generated from subjects' heads from the external noise. Last, head movements were corrected and each subject's data were transformed to a default head position. To remove blink- and pulse-correlated signals from the continuous MEG signals, an independent component analysis (ICA) was performed using the EEGLAB toolbox (Delorme and Makeig, 2004). ICA was performed with 800 maximum steps and 64 PCA dimensions. To capture blink- and pulse-related components, independent component (IC) time series were correlated with vertical EOG and ECG time series. ICs that revealed correlations higher than  $r = 0.4$  and  $z = 3$  were removed and the remaining IC time series were reconstructed. Data were further preprocessed using SPM8 (Well-



**Figure 2.** *A*, Language mask used in the current analysis. The mask consists of bilateral IFG, MTG, STG, SMG, AG, and LATL. *B*, Diagram depicting how the data RDMs are correlated with the model RDMs for each searchlight. Note that the data RDMs change at every time point, whereas the model RDMs remain static. This procedure was repeated for every searchlight and model RDM.

come Trust Centre for Neuroimaging, University College London; r4667) and band-pass filtered between 0.5 and 40 Hz using a fifth-order Butterworth filter.

All of the real word trials were aligned by their UP and segmented into epochs of 1200 ms duration, with a 200 ms silent baseline period before each word. Because our goal was to determine the spatiotemporal dynamics of lexical and semantic activation and competition processes over time, both before and after the UP, epochs were centered on the UP for each word so that the epoch extended from 700 ms before the UP to 500 ms after it. The average baseline period (−700 to −500 ms) was used for baseline correction of each trial. Further, trials contaminated by motion-related artifacts were removed. On average, 3.43 trials were removed due to artifacts ( $SD = 4.22$ ). Finally, time series were downsampled to 250 Hz.

Each subject's data were source localized using both magnetometers and gradiometers. The source space was modeled by a cortical mesh consisting of 8196 vertices. The sensor positions were coregistered to individual subject's T1-weighted MP-RAGE scan using three fiducial points. A single shell model, as implemented in SPM8, was used for forward modeling. Inversion was completed over the entire epoch using the SPM implementation of the minimum norm estimate solution.

ssRSA. ssRSA (Su et al., 2012) involves extraction of data from a searchlight that samples the source estimated cortical space. The searchlights span three spatial dimensions and one temporal dimension, allowing the searchlight to map across the MEG data in both space and time. The spatiotemporal searchlight spheres had a 10 mm spatial radius and a 10 ms temporal radius (i.e., a sliding time window of 20 ms). To construct the data representational dissimilarity matrices for a searchlight sphere centered at a particular point in space and time, the source reconstructed data within the sphere is extracted for every item. The data vectors for each pair of items are compared using correlation distance (i.e.,  $1 - \text{Pearson's } r$ ), which measures the dissimilarity between the brain activity patterns for the pair of items in that spatiotemporal window. This yields a matrix of dissimilarity values (data RDM; see Fig. 3) for the searchlight sphere, where the entries of the matrix correspond to the dissimilarity values of pairs of items.

The data RDMs for each point in space and time are then compared against the dissimilarity values computed for the theoretically relevant models called the model RDMs. Model RDMs were constructed using pairwise dissimilarity values ( $1 - \text{Pearson's } r$ ) for the three theoretical measures of interest that correspond to the trials (Fig. 1*A–C*). LexComp and SemComp were constructed using the behavioral gating data and

were predicted to be sensitive to lexical and semantic competition before the UP. The SemDis model was predicted to tap into target word's unique semantic representations after UP.

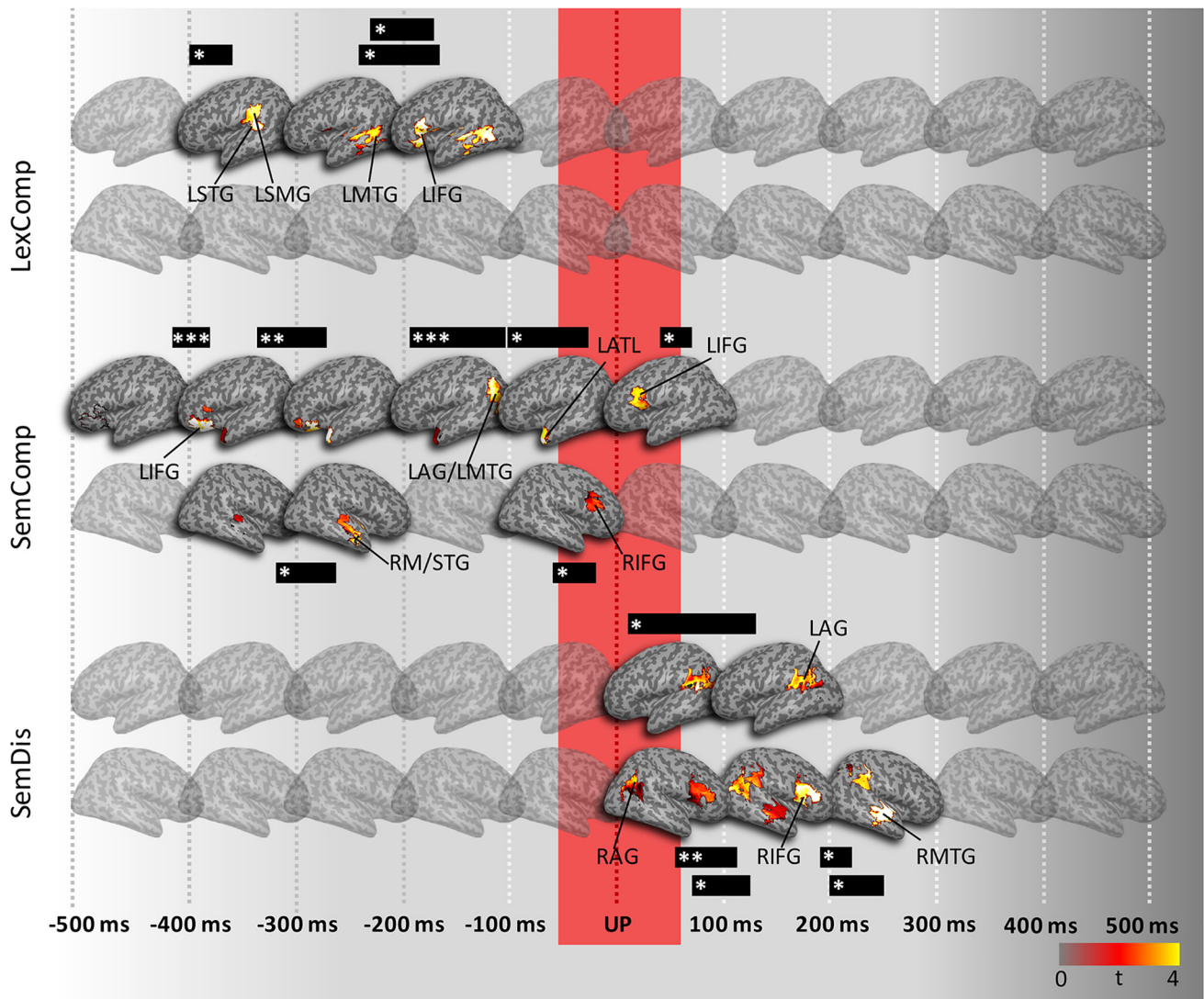
The analysis was restricted to those cortical regions that are consistently reported in studies investigating lexical and semantic processing during language comprehension (Vigneau et al., 2006; Price, 2012; Fig. 2*A*), which included bilateral inferior frontal gyrus (IFG), superior temporal gyrus (STG), middle temporal gyrus (MTG), supramarginal gyrus (SMG), and angular gyrus (AG). The spatial definitions of these regions were taken from the Automated Anatomical Labeling Atlas (Tzourio-Mazoyer et al., 2002) and were fused together as a contiguous mask with 1 mm isotropic spacing. For each participant, data RDMs were constructed for searchlight centroids contained in the mask.

*Statistics and correction for multiple comparisons.* The data RDMs were correlated with model RDMs for each participant using Spearman's rank correlation coefficient (Fig. 2*B*). To allow computation of spatiotemporal clusters in 4D space, correlation time courses for each vertex were placed in the participant's cortical mesh. The 4D matrix consisted of the 3 spatial dimensions of the cortical mesh ( $91 \times 109 \times 91$  spatial points, of which 8196 are on the mesh) and 1 temporal dimension (251 time points). The spatial dimensions here correspond to the size of isotropic 2 mm grid used by SPM8. To correct for the large number of data points tested, we performed cluster permutation based one-sample *t* tests with 1000 permutations with  $p = 0.01$  height and  $p = 0.05$  significance threshold (Nichols and Holmes, 2002). Throughout the results, we present the cluster-level-corrected *p*-values.

## Results

### Competition effects before and after the UP

Figure 3 shows corrected *t*-map snapshots aligned to the UP (0 ms). We predicted a transition from competitive processes before the UP to the activation of the unique semantics of the target word after the UP. Before the UP, we found lexical and semantic competition effects, reflecting the early short-lived parallel activation of candidate lexical and semantic representations as speech is processed over time. The LexComp model showed early transient effects in left STG (LSTG) and left SMG (LSMG) from −400 to −376 ms before the UP ( $p = 0.023$ ) (Fig. 3) and later more sustained effects in LMTG between −224 to −180 ms ( $p = 0.031$ ) and left IFG (LIFG) (BA 45/47) from −244 to −172 ms ( $p =$



**Figure 3.** ssRSA results displaying the corrected  $t$ -maps every 100 ms.  $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ . UP is marked by the red border.

0.04) before the UP. The early effects in LSTG/LSMG reflect the acoustic phonetic computation of spectrotemporal features within speech (Hickok and Poeppel, 2007; Mesgarani et al., 2014). These computations are mapped rapidly onto lexical representations involving the LMTG, which engages later competitive processes between word candidates involving LIFG (Zhuang et al., 2014).

Semantic competition effects captured by the SemComp model showed similarly early effects before the UP starting at approximately the same time as the LexComp model. These early effects were in the LIFG (BA 47) from  $-420$  to  $-392$  ms ( $p = 0.0009$ ) and  $-340$  to  $-288$  ms before the UP ( $p = 0.005$ ) (Fig. 3). The SemComp model revealed short-lived sensitivity in the RSTG/MTG from  $-332$  to  $-292$  ms ( $p = 0.029$ ) before the UP and a sustained effect in the left AG (LAG) and LMTG from  $-96$  to  $-100$  ms ( $p = 0.0009$ ). There were transient effects in the left anterior temporal region (LATL) ( $p = 0.036$ ) from  $-100$  to  $-80$  ms, in right IFG (RIFG) (BA 44/45;  $p = 0.04$ ) from  $-88$  to  $-52$  ms, and in LIFG (BA 44/45;  $p = 0.029$ ) from 44 to 64 ms. These results show that the initial activation of low-level lexical auditory representations gives rise to later semantic representations as activity shifts from STG to LAG, LATL, and bilateral MTG. The

involvement of the LIFG differs in the two competition models. Whereas, in the LexComp model, the LIFG, primarily BA 45, is involved in later competitive processes as the UP approaches, in the SemComp model, BA 47 is initially engaged and shows early and sustained involvement until the UP approaches. Around this point, when sufficient evidence has accrued to identify the target word and activate its unique semantics, we see the involvement of BA 45, which likely reflects later selection processes rather than processes of competition (Zhuang et al., 2014).

#### Transition from semantic competition to target word semantics

Around the UP, when the accumulating speech input enables the target word to be differentiated from the other members of its cohort, there were significant effects of the SemDis model that captured the ease of a target word's feature integration and therefore access to its unique semantic representation. The results of the SemDis model revealed significant model fits over time in bilateral AG, RMTG, and RIFG. The initial effect of the SemDis centered in the L inferior parietal region (Fig. 3), with the cluster including LAG, LSMG, and left posterior MTG ( $p = 0.016$ ) and showing a sustained effect 8–112 ms after the UP. In the right

hemisphere, there were parallel effects in the RIFG from 48–116 ms ( $p = 0.003$ ) and significant clusters in RAG from 72–112 ms ( $p = 0.039$ ) and from 200–248 ms ( $p = 0.025$ ) and in RS/MTG from 192–216 ms ( $p = 0.036$ ). The SemDis results indicate that activity after the UP reflects a process of individuation between the target concept and its cohort neighbors. Furthermore, the overlapping effects of the SemComp and SemDis models in LAG confirm the prominent role of LAG in the conceptual representations and show that UP marks a transition point between the lexical–semantic competitive processes and selection between cohort candidates to the boosting of semantic activation and rapid access to the target’s semantic representation.

## Discussion

This study investigated the spatiotemporal dynamics of the neural computations involved in the transition from the speech input to the target word’s semantic representation by combining cognitive models of competition and semantics with ssRSA analysis of MEG data collected as participants listened to spoken words. Our predictions were based on lexicalist models of speech processing that assume a fully parallel recognition process in which all word candidates that initially fit the speech input are activated and continuously assessed against the accumulating input and those that do not continue to match drop out (Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Gaskell and Marslen-Wilson, 1997). We predicted that, as listeners hear a spoken word, they partially activate lexical and semantic representations of words that match the speech input. When the target word becomes maximally differentiated from its cohort competitors (at the UP), we expected to see a reduction in competitive processes and a boost in the semantic representation of the target word. Consistent with this model of speech comprehension, we found semantic and lexical competitive processes dominated before the UP, which we defined empirically in a gating study as the earliest point in the speech sequence at which only the target word continues to match the sensory input. After the UP, which marks a transition from processes of lexical and semantic competition to a boost in activation of the target word semantics, we found no evidence for competitive processes and selective activation of the semantics of the target word.

This pattern of results is consistent with the view that speech sounds activate meaning representations by means of an optimally efficient language processing system as originally proposed by Marslen-Wilson (1984) and later instantiated in the distributed cohort model of spoken word recognition (Gaskell and Marslen-Wilson, 1997). This optimally efficient system results in a spoken word being recognized as soon as the information becomes available in the speech stream that differentiates it from its competitors (the UP) and accounts for the remarkable earliness with which spoken words can be recognized. This early recognition has important consequences for the wider language system in that the early activation of the semantic and syntactic properties of the word makes them available to be integrated into the developing utterance and discourse representation (Marslen-Wilson and Zwitserlood, 1989), enabling rapid updating of the developing utterance and facilitating rapid and effective communication.

This system is optimally efficient in the sense that it makes maximally effective use of incoming sensory information to guide dynamic perceptual decisions. Although behavioral studies have supported this account (Gaskell and Marslen-Wilson, 2002; Apfelbaum et al., 2011), the present study is the first to reveal the neural dynamics associated with this optimally efficient system.

## Lexical access and competition

The MEG data show that processing a spoken word is initially dominated by processes of lexical and semantic competition, reflecting the parallel activation of multiple phonologically based word candidates that initially match the speech input and decay as the UP approaches. Our measure of lexical competition was based on cohort statistics, which reflect the structure and organization of the mental lexicon (Marslen-Wilson, 1987; Cibelli et al., 2015). Lexical competition effects were observed well before the UP and initially engaged LSTG/LSMG, regions that are involved in acoustic and phonological analysis (Rauschecker and Scott, 2009). These effects subsequently spread to the LMTG, which is associated with processing the semantic properties of words (Binder et al., 2009), and were followed by effects in BA 45, which, in previous studies using similar manipulations, has been shown to be responsive to selecting between activated cohort competitors (Zhuang et al., 2014).

There were similarly early effects of semantic competition resulting from the activation of the semantics of the multiple word candidates initiated by the activation of multiple lexical candidates. These early effects of semantic competition involved the LIFG (BA 47) and spread into the LATL, followed by effects in the LAG and then the ATL again as the UP neared and the number of semantic competitors declined. The LAG is frequently associated with a variety of semantic-based processes (Binder et al., 2009) such as semantic feature integration (Price et al., 2016), whereas the LATL is thought to be a core region within the amodal semantic hub (Patterson et al., 2007). The finding that the effects of semantic competition flow between these regions over time suggests their joint sensitivity to semantic effects. Finally, around the UP, we found effects of the SemComp model in bilateral IFG (BA 45), most likely reflecting the resolution of semantic competition and selection of the appropriate candidate (Grindrod et al., 2008).

Although the effects of cohort competition were seen at both the lexical and semantic levels, semantic competition effects were more robust and sustained than lexical competitive effects. The SemComp model showed initial early effects primarily in LIFG BA 47, a region associated with cohort competition effects in previous studies of spoken word recognition (Zhuang et al., 2014) and later effects in more superior inferior frontal regions (BA 44/45), which, in previous studies (Zhuang et al., 2014), has been shown to be involved in cohort selection processes. These results suggest that, whereas a word’s phonological form carries its meaning, processes of cohort competition and selection mainly involve the semantic representations of cohort competitors (Gaskell and Marslen-Wilson, 1999; Devereux et al., 2016).

After the UP, there were no competitive effects of either phonology or semantics, supporting the claim that the UP marks the earliest point at which a word can be reliably differentiated from its cohort competitors (Marslen-Wilson and Tyler, 1980; Tyler, 1984). Consistent with an optimally efficient speech recognition system, after the UP, we only found effects of the target word semantics, suggesting that, before the UP, the semantics of the target word were not differentiated from the semantics of the other words in the cohort.

## Rapid access to target word semantics

During the process of recognizing a spoken word, the target word’s semantic representation is partially activated, along with the partial activation of the semantics of all of the other members of the cohort. We predicted that, around the UP, after the resolution of cohort competition, the partial activation of a target

word's semantic representation would be boosted because the speech input at that point would enable a unique phonology-to-semantic mapping and the semantics of the competitor words would have decayed. The SemDis measure, based on a well validated feature-based model of lexical semantics, captures the relative ease with which the semantic features of a concept can be integrated (Taylor et al., 2012). Consistent with an optimally efficient speech recognition system, soon after the UP, we found effects of the target word semantics, suggesting that, before the UP, the semantics of the target word were not sufficiently differentiated from the semantics of the other words in the cohort. The SemDis model revealed a number of regions (bilateral AG, MTG) that have been associated previously with lexical semantic information. Moreover, the AG has been claimed to function as an integrative hub in which different kinds of semantic feature information can be combined (Seghier, 2013; Bonnici et al., 2016; Price et al., 2016), thus providing a potential neurobiological substrate for the integration of features into a holistic concept.

The results of the SemComp and the SemDis models suggest that the activation of lexical semantic information is rapid. The SemComp model that incorporated the target word's semantic representation showed retrieval effects starting from  $-244$  ms before the UP, suggesting that the target word's semantic representation is partially activated before the UP, along with the semantics of the cohort competitors.

The SemDis model, in contrast, started showing effects 8 ms after the UP. The absence of effects of the SemDis model before the UP suggests that the boost of target word's semantic activation depends on the unique identification of the spoken word. Although previous studies have shown that meaningful words are distinguished as early as 50 ms after the UP (MacGregor et al., 2012) and a word's semantic representation is activated within 200 ms (Pulvermüller et al., 2005), we found even earlier effects in the current study, perhaps due to higher spatiotemporal sensitivity achieved by ssRSA compared with a more conventional event related potential/field approach. The latter involves averaging the signal across all channels, which results in losing the spatial information embedded in the signal.

## Conclusion

Our findings suggest that the sound-to-meaning mapping during natural speech comprehension involves rapid dynamic computations aligned around the UP. By testing quantifiable cognitive models of key cognitive processes as they occur in real time in the brain, this research revealed the dynamic cortical networks that relate to partially activated lexical and semantic representations of the cohort candidates before the UP and the gradual resolution of competitive processes as speech accrues over time, the pool of candidate words narrows down to a single concept, and the target word's unique conceptual representation is boosted.

## References

- Apfelbaum KS, Blumstein SE, McMurray B (2011) Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychon Bull Rev* 18:141–149. [CrossRef Medline](#)
- Baayen R, Piepenbrock R, van Rijn H (1993) The CELEX lexical data base on CD-Rom. Linguistic Data Consortium. Release 1 CD-ROM. University of Pennsylvania, Philadelphia.
- Baroni M, Lenci A (2010) Distributional memory: a general framework for corpus-based semantics. *Computational Linguistics* 36:673–721. [CrossRef](#)
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796. [CrossRef Medline](#)
- Bonnici HM, Richter FR, Yazar Y, Simons JS (2016) Multimodal feature

- integration in the angular gyrus during episodic and semantic retrieval. *J Neurosci* 36:5462–5471. [CrossRef Medline](#)
- Cibelli ES, Leonard MK, Johnson K, Chang EF (2015) The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. *Brain Lang* 147:66–75. [CrossRef Medline](#)
- Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV (2012) The representation of biological classes in the human brain. *J Neurosci* 32:2608–2618. [CrossRef Medline](#)
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21. [CrossRef Medline](#)
- Devereux BJ, Tyler LK, Geertzen J, Randall B (2014) The Centre for Speech, Language and the Brain (CSLB) concept property norms. *Behav Res Methods* 46:1119–1127. [CrossRef Medline](#)
- Devereux BJ, Taylor KI, Randall B, Geertzen J, Tyler LK (2016) Feature statistics modulate the activation of meaning during spoken word processing. *Cogn Sci* 40:325–350. [CrossRef Medline](#)
- Gaskell MG, Marslen-Wilson WD (1997) Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes* 12:613–656. [CrossRef](#)
- Gaskell MG, Marslen-Wilson WD (1999) Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science* 23:439–462. [CrossRef](#)
- Gaskell MG, Marslen-Wilson WD (2002) Representation and competition in the perception of spoken words. *Cognit Psychol* 45:220–266. [CrossRef Medline](#)
- Grindrod CM, Bilenko NY, Myers EB, Blumstein SE (2008) The role of the left inferior frontal gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Res* 1229:167–178. [CrossRef Medline](#)
- Grosjean F (1980) Spoken word recognition processes and the gating paradigm. *Percept Psychophys* 28:267–283. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Kanerva P, Kristoferson J, Holst A (2000) Random indexing of text samples for Latent Semantic Analysis. In: LR Gleitman, AK Josh (eds.), *Proc 22nd Annual Conference of the Cognitive Science Society (U Pennsylvania)* p. 1036. Mahwah, New Jersey: Erlbaum.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- MacGregor LJ, Pulvermüller F, van Casteren M, Shtyrov Y (2012) Ultra-rapid access to words in the brain. *Nat Commun* 3:711. [CrossRef](#)
- Marslen-Wilson W, Zwitserlood P (1989) Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance* 15:576–585. [CrossRef](#)
- Marslen-Wilson WD (1984) Function and process in spoken word recognition: a tutorial review. *Atten Perform* 10:125–150.
- Marslen-Wilson WD (1987) Functional parallelism in spoken word-recognition. *Cognition* 25:71–102. [CrossRef Medline](#)
- Marslen-Wilson W, Tyler LK (1980) The temporal structure of spoken language understanding. *Cognition* 8:1–71. [Medline](#)
- Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. Available from: <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ184160> (Accessed May 10, 2011).
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cognit Psychol* 18:1–86. [CrossRef Medline](#)
- McRae K, Cree GS, Seidenberg MS, McNorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. *Behav Res Methods* 37:547–559. [CrossRef Medline](#)
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343:1006–1010. [CrossRef Medline](#)
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25. [CrossRef Medline](#)
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976–987. [CrossRef Medline](#)
- Price AR, Peelle JE, Bonner MF, Grossman M, Hamilton RH (2016) Causal evidence for a mechanism of semantic integration in the angular gyrus as

- revealed by high-definition transcranial direct current stimulation. *J Neurosci* 36:3829–3838. [CrossRef Medline](#)
- Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62:816–847. [CrossRef Medline](#)
- Pulvermüller F, Shtyrov Y, Ilmoniemi R (2005) Brain signatures of meaning access in action word recognition. *J Cogn Neurosci* 17:884–892. [CrossRef Medline](#)
- Randall B, Moss HE, Rodd JM, Greer M, Tyler LK (2004) Distinctiveness and correlation in conceptual structure: behavioral and computational studies. *J Exp Psychol Learn Mem Cogn* 30:393–406. [CrossRef Medline](#)
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724. [CrossRef Medline](#)
- Seghier ML (2013) The angular gyrus multiple functions and multiple subdivisions. *Neuroscientist* 19:43–61. [CrossRef Medline](#)
- Shannon CE (2001) A mathematical theory of communication. *ACM SIGMOBILE Mob Comput Commun Rev* 5:3–55.
- Shannon CE, Weaver W (1949) *The mathematical theory of communication*. Urbana-Champaign, IL: University of Illinois.
- Su L, Fonteneau E, Marslen-Wilson W, Kriegeskorte N (2012) Spatiotemporal searchlight representational similarity analysis in EMEG source space. *International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, pp 97–100. [CrossRef](#)
- Taylor KI, Devereux BJ, Acres K, Randall B, Tyler LK (2012) Contrasting effects of feature-based statistics on the categorisation and basic-level identification of visual objects. *Cognition* 122:363–374. [CrossRef Medline](#)
- Tyler LK (1984) The structure of the initial cohort: evidence from gating. *Percept Psychophys* 36:417–427. [CrossRef Medline](#)
- Tyler LK, Wessels J (1985) Is gating an on-line task? Evidence from naming latency data. *Percept Psychophys* 38:217–222. [CrossRef Medline](#)
- Tyler LK, Chiu S, Zhuang J, Randall B, Devereux BJ, Wright P, Clarke A, Taylor KI (2013) Objects and categories: feature statistics and object processing in the ventral stream. *J Cogn Neurosci* 25:1723–1735. [CrossRef Medline](#)
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273–289. [CrossRef Medline](#)
- Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432. [CrossRef Medline](#)
- Warren P, Marslen-Wilson W (1987) Continuous uptake of acoustic cues in spoken word recognition. *Percept Psychophys* 41:262–275. [CrossRef Medline](#)
- Warren P, Marslen-Wilson W (1988) Cues to lexical choice: discriminating place and voice. *Percept Psychophys* 43:21–30. [CrossRef Medline](#)
- Zhuang J, Tyler LK, Randall B, Stamatakis EA, Marslen-Wilson WD (2014) Optimally efficient neural systems for processing spoken language. *Cereb Cortex* 24:908–918. [CrossRef Medline](#)