



## Full Length Article

The anterior left inferior frontal gyrus contributes to semantic unification<sup>☆</sup>Zude Zhu<sup>a</sup>, Peter Hagoort<sup>b,c</sup>, John X. Zhang<sup>d</sup>, Gangyi Feng<sup>a</sup>, Hsuan-Chih Chen<sup>d</sup>, Marcel Bastiaansen<sup>b,c</sup>, Suiping Wang<sup>a,\*</sup><sup>a</sup> Center for the Study of Applied Psychology, South China Normal University, China<sup>b</sup> Radboud University Nijmegen, Donders Institute, Centre for Cognitive Neuroimaging, The Netherlands<sup>c</sup> Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands<sup>d</sup> Department of Psychology, Chinese University of Hong Kong, Hong Kong

## ARTICLE INFO

## Article history:

Accepted 15 February 2012

Available online 23 February 2012

## Keywords:

Semantic unification

aLIFG

Violation

## ABSTRACT

Semantic unification, the process by which small blocks of semantic information are combined into a coherent utterance, has been studied with various types of tasks. However, whether the brain activations reported in these studies are attributed to semantic unification *per se* or to other task-induced concomitant processes still remains unclear. The neural basis for semantic unification in sentence comprehension was examined using event-related potentials (ERP) and functional Magnetic Resonance Imaging (fMRI). The semantic unification load was manipulated by varying the goodness of fit between a critical word and its preceding context (in high cloze, low cloze and violation sentences). The sentences were presented in a serial visual presentation mode. The participants were asked to perform one of three tasks: semantic congruency judgment (SEM), silent reading for comprehension (READ), or font size judgment (FONT), in separate sessions. The ERP results showed a similar N400 amplitude modulation by the semantic unification load across all of the three tasks. The brain activations associated with the semantic unification load were found in the anterior left inferior frontal gyrus (aLIFG) in the FONT task and in a widespread set of regions in the other two tasks. These results suggest that the aLIFG activation reflects a semantic unification, which is different from other brain activations that may reflect task-specific strategic processing.

© 2012 Elsevier Inc. All rights reserved.

## Introduction

An important goal of language comprehension is to understand the meaning from the given information flow. Semantic unification or the process by which small pieces of word-level information are combined into larger message-level representations (Hagoort et al., 2009) is crucial to this goal. Neuroimaging studies using Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI) have revealed a number of brain regions involved in semantic unification, such as the left inferior frontal gyrus (LIFG) (Braze et al., 2011; Diaz et al., 2011; Hagoort et al., 2009; Huang et al., 2012; Rodd et al., 2005; Tesink et al., 2009; Zhu et al., 2009), anterior temporal lobe (ATL) (Crinion et al., 2003; Crinion and Price, 2005; Warren et al., 2009), posterior superior temporal gyrus (pSTG) (Friederici et al., 2009; Friederici et al., 2003) and angular gyrus (AG) (Friederici et al., 2003, 2009).

Recent reviews (Hagoort et al., 2009; Zhu et al., 2011) describe the LIFG as often being activated during semantic unification in fMRI studies. However, some experiments have reported activation in the

temporal cortex during semantic unification tasks, but no or weak activation in the LIFG (Crinion et al., 2003; Friederici et al., 2003; Price, 2010; Rogalsky and Hickok, 2009; Vandenberghe et al., 2002). Such discrepancies may result from differences in the methodology and design of previous studies. First, contrasts adopted in the literature to reveal semantic processing may be confounded by other factors. For example, a stronger activation in processing incongruent compared with congruent sentences was usually employed to localize brain regions for semantic unification. However, compared with congruent sentences, the incongruent sentences captured more attention and triggered an automatic semantic recovery, a process that was not necessarily engaged during the reading of congruent sentences (Kaan and Swaab, 2003; Zhu et al., 2009). Second, different tasks across different studies may also induce processing due to task-specific strategies. For example, a comprehension probe task (Newman et al., 2009) not only requires semantic unification, but also taxes additional working memory resources (Caplan et al., 2008), which may conflict with the effects of interest (Newman et al., 2009). In the violation detection task (Ni et al., 2000), semantic processing may terminate earlier in the incongruent condition than in the congruent condition due to a specific reading strategy (Huang et al., 2012; Zhu et al., 2009). Although activation in the LIFG has been reported in a wide range of tasks, including semantic congruency judgment (Constable et al., 2004; Rueschemeyer et al., 2006; Zhu et al., 2009), meaningfulness

<sup>☆</sup> The authors declare no conflict of interest.

\* Corresponding author at: Center for the Study of Applied Psychology, South China Normal University, Guangzhou, 510631, China. Fax: +86 20 8521 6033.

E-mail address: [suiping@scnu.edu.cn](mailto:suiping@scnu.edu.cn) (S. Wang).

rating (Humphries et al., 2007), reading for comprehension (Hagoort et al., 2004; Tesink et al., 2009), comprehension probe test (Just et al., 1996; Mason and Just, 2007; Newman et al., 2009; Ye and Zhou, 2009), priming (Devauchelle et al., 2009), and violation detection (Ni et al., 2000), it has been suggested that these tasks all involve explicit attentional control (Crinion et al., 2003; Van Petten and Luka, 2006). It is possible that such control processes may interact with semantic unification and constitute an alternative explanation for the LIFG activations observed in these studies.

It is therefore necessary to identify brain regions that are responsible for semantic unification as opposed to other cognitive processes that may accompany semantic unification. To this end, we constructed three types of sentences (see Fig. 1), including high cloze (HC), low cloze (LC) and violation sentences (SV) (Hagoort and Brown, 1994). High-cloze sentences were sentences in which a noun at a given position (the critical word) was highly expected. Low-cloze sentences were semantically correct, but the noun at the critical word position was unexpected. In violation sentences, the noun at the critical word position did not semantically fit into the sentence context. As a highly automatic process, semantic unification is assumed to occur for violation sentences even if eventually no meaningful semantic representation can be established. Thus, this means that in terms of the distinction between effort and outcome, semantic unification is more associated with processing effort as opposed to the outcome. In this sense, the high cloze and violation conditions engage the least and most semantic unification effort, respectively. This understanding is consistent with the common interpretation of the N400 effect, which is that enhanced N400 in the violation sentences indicates increased difficulty in semantic unification (Hagoort, 2008; Hagoort et al., 2009; Kuperberg, 2007; but see Kutas and Federmeier, 2011; Van Petten and Luka, 2012). This understanding is also consistent with eye-tracking data and self-paced reading data in which the reading time at the target word linearly increased across the HC, LC and SV sentences (e.g., Wang et al., unpublished data; Wang et al., 2011). However, at the end of the sentence, the reading time was shorter in the incongruent compared with the congruent sentence (e.g., Kuperberg et al., 2003; Zhu et al., 2009), suggesting the failure of semantic unification in the incongruent condition. Thus, semantic unification may terminate early in the SV condition, but there should still be semantic unification effort at the target word presented in the middle of the sentence. In our analyses, the load effect (indicated by a parametric modulation) was time-locked to the critical word, which was located in the middle of the sentence to reflect the semantic unification at the target position in an fMRI study, and was parallel with the load effect analysis in the ERP study (with trend analysis) to a time-lock on the critical words.

Furthermore, we employed three tasks: semantic congruency judgment (SEM), reading for comprehension (READ), and font size judgment (FONT). The first two tasks are often used in the literature and are known to require explicit semantic unification. In the FONT task, participants are oriented to the font size of the words in a sentence and are not required to process the semantic meaning of the

sentence. However, based on the general understanding that semantic unification is a highly automatic process, semantic unification is supposed to occur in this task as well. For example, using the N400 effect as an index, Relander et al. (2009) observed semantic unification in an incidental task that did not require any explicit attention to the semantic content of the linguistic stimuli. Thus we intended to use a conjunction analysis to identify the brain regions associated with semantic unification that was common to the three different tasks, i.e., regions that contributed to the semantic unification should be parametrically modulated by the semantic unification load in all three tasks. To confirm that semantic unification had occurred in the FONT task, a parallel ERP session was also conducted to examine how the N400 component was modulated by sentence type. Such modulation is generally viewed as an electrophysiological signature for semantic unification (Hagoort, 2008).

## Methods

### Participants

Two groups of 27 healthy native Chinese speakers voluntarily participated in the study in exchange for monetary compensation (ERP session, 10 males and 17 females, aged between 18 and 25 years, mean age = 21.5 years; fMRI session, 11 males and 16 females, aged between 18 and 28 years, mean age = 22.3 years). All of the subjects were right-handed with normal or corrected-to-normal vision, and none of the subjects exhibited any neurological disorders. An informed written consent was obtained from each individual following a protocol approved by the local ethics committee.

### Stimuli

As illustrated in Fig. 1, the stimuli consisted of three types of sentences that were high cloze sentences (HC), low cloze sentences (LC), and violation sentences (SV). A base set of 315 sentences with high constraint context was first constructed as the HC sentences. Each sentence was then modified to produce a LC sentence by replacing a selected noun with an unexpected noun, which was semantically congruent with the context. A SV sentence was modified by replacing the critical word with a word incongruent with the context. That is, the three versions of each base sentence differed from each other only by one critical word. Each sentence contained 11 words. The critical word occurred in the 6th position in half of the sentences and in the 7th position in the other half. Across the three conditions, the critical words were matched in frequency (mean and standard deviation, HC:  $24.7 \pm 48.0$ ; LC:  $23.8 \pm 57.1$ ; SV:  $22.1 \pm 51.6$ ,  $F < 1$ ) and in visual complexity (mean stroke number and standard deviation, HC:  $8.2 \pm 2.4$ ; LC:  $7.4 \pm 2.1$ ; SV:  $8.0 \pm 2.2$ ,  $F < 1$ ).

The mean cloze probability of the HC sentences was 63.5% ( $SD = 19.3$ , range = 30–100), as rated by 33 other participants from the same subject pool using the cloze test (Kutas and Hillyard, 1984). During the rating, the initial part of the sentence (up until the critical word) was presented and the participants were asked to continue the sentence with the first noun that came to mind to make the sentence meaningful. The cloze probability of the LC nouns and of the SV nouns was zero.

Semantic acceptability was rated by another 20 new participants from the same subject pool on a 5-point scale, with 1 for extremely unacceptable and 5 for completely acceptable. The average ratings for the HC, LC, and SV sentences were  $4.2 \pm 0.3$ ,  $4.1 \pm 0.3$ , and  $1.7 \pm 0.4$ , respectively. There were significant across-group differences ( $F(2, 618) = 4775.7$ ,  $p < 0.001$ ) and all the pair-wise comparisons were significant ( $p$ -values  $< 0.001$ ).

In the ERP study, a total of 9 lists were constructed, each containing 35 HC sentences, 35 LC sentences, and 35 SV sentences. Within each list, no two sentences were associated with the same base

| Conditions      | Sample sentences   |
|-----------------|--|
| High cloze (HC) | 睡觉前他调了闹钟后放心地睡了<br>He set <u>alarm clock</u> before he went to bed and slept soundly. |
| Low cloze (LC)  | 睡觉前他调了灯光后放心地睡了<br>He set <u>light</u> before he went to bed and slept soundly.       |
| Violation (SV)  | 睡觉前他调了月亮后放心地睡了<br>He set <u>moon</u> before he went to bed and slept soundly.        |

Fig. 1. Sample sentences used in the three conditions with the critical words underlined.

sentence. To prevent explicit attention to the critical word position, 105 incongruent filler sentences were constructed with the violation occurring equally likely at the 8th, 9th or 10th position of the sentence. The sentence length was also 11 words. From these fillers, 35 sentences were pseudo-randomly drawn and added to each of the above 9 lists. Within each list, the different types of sentences were randomly intermixed to balance their presentation order.

In the fMRI study, the stimuli were similar with those used in the ERP study, except that the number of sentences was reduced to 24 for each condition and 18 filler sentences were on each list due to time constraints.

### Tasks

The E-Prime software package (Psychology software Tool, Pittsburgh, PA) was used for the stimulus presentation and response collection. For each trial, the subjects were asked to fixate on to a 500-ms presentation of an asterisk in the center of the screen, and then one sentence was presented word by word. Each word was shown for 300 ms and was followed by a 300-ms blank screen before the next word was displayed. The offset of the final word completed the sentence presentation.

For the reading task, the participants were asked to read the sentence for comprehension, but no overt response was required. They were also asked to maintain fixation on a cross sign for 4 s following the sentence presentation. For the semantic task, after a sentence was presented, the participants were required to indicate whether it was semantically congruent or incongruent as fast and as accurately as possible within a 4-s response window by pressing a button. For the font size task, a cross fixation would follow the sentence presentation, replaced 500 ms later by a Chinese word (the Chinese word pinyin is 'Ceshi', which means 'TEST'). The participants were required to judge whether this word was of the same font size as the words in the preceding sentence by pressing a button, which was also within a 4-s response window. The sentences were always presented in size 40 Times New Roman, and the test word was of the same font type but was either larger or smaller in size (50 or 34). The font sizes were determined based on pilot studies for the appropriate level of task difficulty to avoid ceiling or floor effects. When the responses were required, both the response accuracy and the response time were collected. The responses made after the 4-s limits were treated as errors. The response keys were counterbalanced across the subjects.

Each participant completed three sessions, one for each task. The reading task was always presented first, while the other two sessions counterbalanced in the task order. This was to prevent contamination of the reading task, should it follow the semantic or font size task. For example, participants may try to evaluate semantic congruency or attend to font size even if not required for the reading task. Each session involved the use of 3 lists of sentences. The 9 lists described earlier were pseudo-randomly assigned so that each list was equally likely to appear in all three of the tasks, and no two sentences in one session shared the same base sentence. The participants received 10 trials of practice using extra materials before each session. Other than the reading task, the subjects were provided with feedback on their responses to make sure that they understood the task. However, no feedback was provided for any task during the testing sessions.

### ERP study

For the behavioral data, the accuracy and reaction time (RT) were collected by the E-Prime software package. The mean accuracy and RT for the different conditions in the semantic and font size tasks were entered into a repeated-measures analysis of variance (ANOVA) using SPSS 13.

The EEG data were collected from a 29 standard channel (10/20 system) ActiCap (Brain Products, Munich, Germany), in addition to the three electrodes used to measure eye movement, as shown in

**Fig. 2.** The EEG was continuously recorded with a 125 Hz low-pass filter sampled at 500 Hz. All of the electrode impedances were kept below 10 k $\Omega$  during the recording. The EEG data were re-referenced offline to the average of the two mastoids and then filtered with a 0.5–25 Hz bandpass filter. The critical epochs ranged from –200 ms to 1200 ms relative to the onset of the critical word, with –200 ms to 0 ms serving as the baseline. The artifact rejection criterion was  $\pm 90$   $\mu$ V. The epochs that included the eye blinks were removed by visual inspection. The kept trial for the HC, LC, SV was  $33 \pm 2.3$ ,  $33 \pm 2.3$ ,  $33 \pm 2.2$  in the FONT task,  $32 \pm 3.3$ ,  $33 \pm 2.6$ ,  $33 \pm 2.5$  in the SEM task, and  $31 \pm 4.0$ ,  $32 \pm 3.8$ ,  $32 \pm 3.9$  in the READ task, without significant differences among the sentence types ( $F(2, 52) = 1.8$ ,  $p > .2$ ), task ( $F < 1$ ), and the sentence type by task interaction ( $F(4, 104) = 1.3$ ,  $p > .2$ ).

The critical time window was 300–500 ms for the N400, but the N400 effect was often followed by a late positive component (Van Petten and Luka, 2012). Thus, the 500–1200 ms time window would also be analyzed and discussed. Therefore, for each participant, the kept trials were averaged into three segments (300–500, 500–800 and 800–1200) for each condition separately. Next, the values of the five electrodes were averaged in four regions: left anterior (including Fp1, F3, F7, FC1 and FC5), left posterior (including CP1, CP5, P3, P7 and O1), right anterior (including Fp2, F4, F8, FC2 and FC6), and right posterior (including CP2, CP6, P4, P8 and O2). Then, a repeated-measures ANOVA was performed on the averaged data, with the following factors: sentence type (HC, LC, SV), task (FONT, READ, SEM), hemisphere (Left, Right), and region (Anterior, Posterior). To better capture the modulation effect on N400, we further performed a trend analysis of the N400 amplitude across the three conditions in the posterior region. Planned comparisons among the conditions were also performed to detect significant interactions involving at least one condition factor. The Greenhouse–Geisser correction was applied when evaluating effects with more than one degree of freedom to protect against Type I errors from sphericity violations.

### fMRI study

#### Imaging procedure

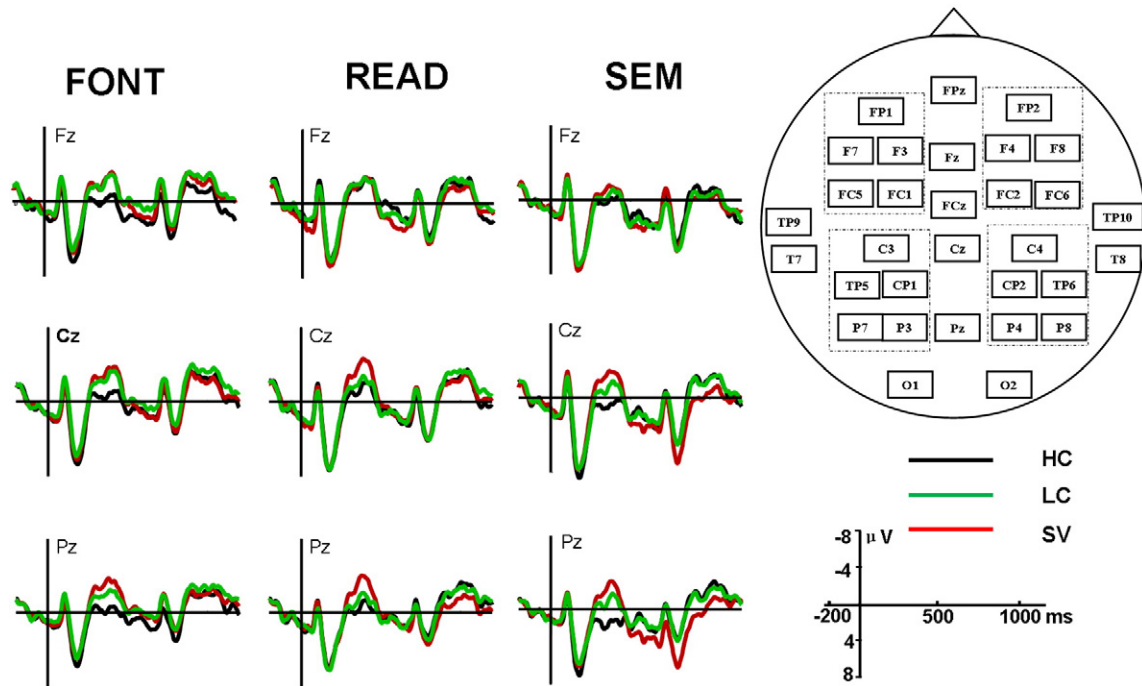
The participants lay in a supine position inside of the scanner, wearing MRI-compatible earphones and goggles (Resonance Technology Company, Los Angeles, CA) and holding a button box. They were told not to move their head (restrained by padding) inside of the scanner and that they could close their eyes for a short rest between the two successive runs. Each run had lasted for 19 min and 54 s, including two self-paced rests, which were usually approximately 1 min.

#### fMRI data acquisition

The data acquisition was performed using a Siemens 1.5 T MR machine. Interleaved whole-brain echo-planar images (EPIs) were acquired with ascending slice order (TR = 2000 ms, TE = 43 ms, flip angle = 90°, 26 slices, slice thickness = 5 mm, voxel size = 3.6  $\times$  3.6  $\times$  5 mm<sup>3</sup>). Following the functional runs, a high resolution T1-weighted scan was acquired using a SAG sequence (192 slices, TE = 3.03 ms; slice thickness = 1 mm; voxel size = 1  $\times$  1  $\times$  1 mm<sup>3</sup>).

#### Data analysis

The data were analyzed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/-spm8/>). The preprocessing included motion correction by means of a rigid body registration along 3 rotations and 3 translations, correction of the slice acquisition time (in reference to a middle 14th slice), co-registration between the EPI and structural images, normalization to a standard MNI T1 template, interpolation to 2  $\times$  2  $\times$  2 mm<sup>3</sup> voxel size, high-pass filtering with a time constant of 128 s, and spatial smoothing with an 8-mm FWHM Gaussian kernel. The statistical analysis was performed using general linear modeling (GLM) in the first level or individual analysis and the random effect models in the second level or group analysis.



**Fig. 2.** Scheme for the electrodes and the N400 effect. The ERP waveforms on the Fz, Cz and Pz in FONT, READ and SEM were shown. The HC for the high cloze sentence, LC for the low cloze sentence, and SV for the violation sentence. In the schematic of montage, four areas (left anterior, right anterior, left posterior and right posterior) corresponding to the region and hemispheric factors in data analysis were indicated by dotted line.

Note that parametric modulation analysis rather than pairwise contrasts was chosen for GLM because specific response strategies may complicate results from the latter. For example, the LC may trigger inhibitory function to suppress the expected words, whereas the SV may trigger a violation detection when the readers encounter the critical words. Such strategies may not be constant across the different contrasts. Thus, a parametric modulation intended to capture the monotonic change across the conditions would be more suitable for the present research goal, that is, to localize the regions modulated by the semantic unification load. Furthermore, the parametric modulation may also help to reduce Type I errors associated with multiple pairwise comparisons.

For each task, the GLMs were first performed to estimate the parametric modulation effect based on the sentence type. Because the N400 amplitude was parametrically modulated (see ERP results section), as often reported in the literature, we used the values,  $-1$ ,  $0$ , and  $1$ , in the design matrix for the HC, LC and SV to mirror the monotonic trend across the conditions. The event of interest was set at the onset of the critical word with a  $0.6$  s duration (Hagoort et al., 2004). All of the other words, self-paced breaks, and 6 motion parameters were modeled as events of no interest. A conjunction analysis was then performed to reveal common activations across the three tasks and to illustrate the detailed activation patterns in the identified regions. All of the analyses had used an uncorrected voxel-level threshold of  $0.01$  combined with a cluster-level threshold of  $0.05$  to correct for multiple comparisons, unless otherwise specified. The brain region coordinates were reported in the MNI space and labeled with Brodmann's areas.

## Results

### Behavioral results

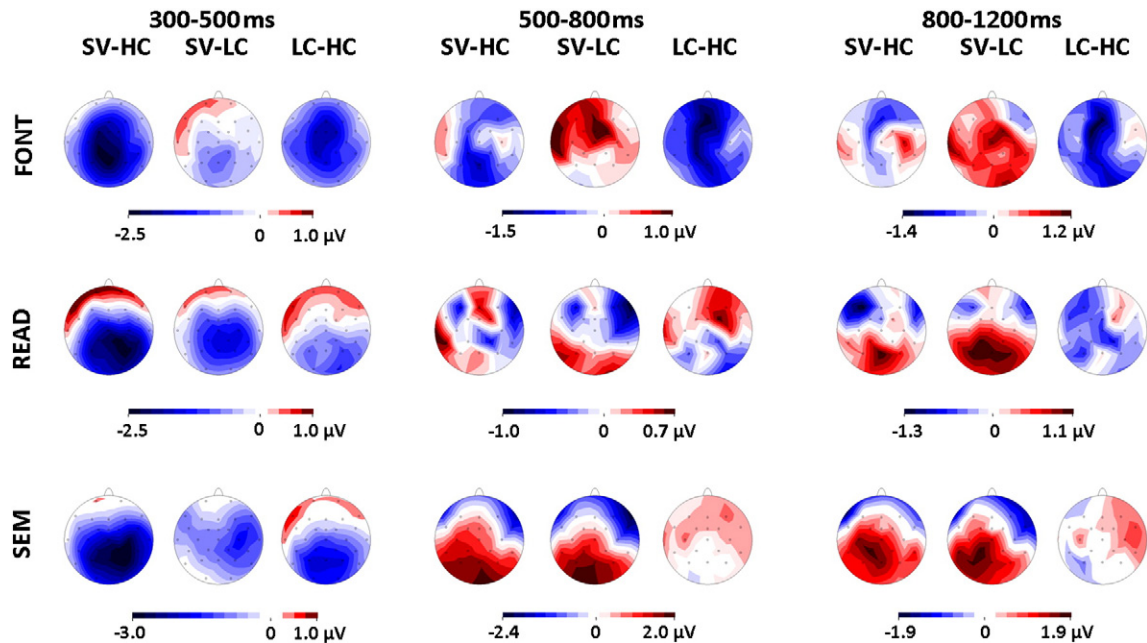
In the ERP session, for the SEM task, the mean accuracies (with SD) were  $97\% \pm 4.2$ ,  $90.5\% \pm 12.4$ , and  $94.7\% \pm 5.3$  for the HC, LC, and SV conditions, respectively. The main effect of sentence type was close to significance ( $F(2, 52) = 5.60$ ,  $p = 0.06$ ). The pairwise comparisons showed significant differences between the HC and LC

( $p < 0.01$ ), but not between the HC and SV ( $p = 0.22$ ), nor between the LC and SV ( $p = 0.40$ ), with Bonferroni correction. The corresponding mean RTs were  $674 \pm 189$ ,  $723 \pm 221$ , and  $645 \pm 206$  ms for HC, LC and SV, respectively. The main effect of sentence type was significant ( $F(2, 52) = 3.54$ ,  $p < 0.05$ ). The pairwise comparisons showed significant differences between the LC and SV ( $p < 0.05$ ), but not between the HC and SV ( $p = 0.98$ ), nor between the HC and LC ( $p = 0.25$ ), with Bonferroni correction. For the FONT task, the mean accuracies were  $94.3 \pm 12.1$ ,  $93.9 \pm 13.3$ , and  $94.1 \pm 8.8$  for the HC, LC and SV conditions, respectively. The corresponding mean RTs were  $963 \pm 210$ ,  $970 \pm 202$ , and  $938 \pm 189$  ms. Moreover, no significant effects were found with accuracy ( $F < 1$ ) nor with RTs ( $F(2, 52) = 1.34$ ,  $p = .27$ ).

In the fMRI session, for the SEM task, the mean accuracies were  $93.3\% \pm 8.3$ ,  $89.3\% \pm 7.7$ , and  $93.7\% \pm 4.9$  for the HC, LC, and SV conditions, respectively. The main effect of the sentence type was significant ( $F(2, 52) = 3.59$ ,  $p < 0.05$ ). The pairwise comparisons showed significant differences between the HC and LC ( $F(1, 26) = 5.14$ ,  $p < 0.05$ ), between the LC and SV ( $F(1, 26) = 6.74$ ,  $p < 0.01$ ), but not between the HC and SV ( $F < 1$ ). The corresponding mean RTs were  $829 \pm 409$ ,  $884 \pm 500$ , and  $637 \pm 366$  ms. The main effect of sentence type was also significant ( $F(2, 52) = 17.58$ ,  $p < 0.001$ ) and the pairwise comparisons showed significant differences between the HC and SV ( $F(1, 26) = 20.19$ ,  $p < 0.001$ ), between the LC and SV ( $F(1, 26) = 20.08$ ,  $p < 0.001$ ), and a borderline significant difference between the HC and LC ( $F(1, 26) = 3.44$ ,  $p < 0.08$ ). For the FONT task, the mean accuracies were  $80.4 \pm 18.6$ ,  $82.3 \pm 19.9$ , and  $83.9 \pm 16.8$  for the HC, LC and SV conditions, respectively. The corresponding mean RTs were  $1002 \pm 271$ ,  $1004 \pm 262$ , and  $999 \pm 296$  ms. No significant effects were found with either measure ( $p > 0.2$ ).

### ERP results

The aim of this part was to confirm that semantic unification occurred in all of the tasks. As shown in Figs. 2 and 3, the manipulation of the semantic unification load was successful as indicated by the N400 amplitude (300–500 ms) differences across the three sentence



**Fig. 3.** Topographic maps for the HC, LC, SV pairwise comparisons. There were three segments (300–500, 800–800 and 800–1200 ms) in each comparison and in each task (FONT, READ and SEM).

types in all of the three tasks, with a significant linear enhanced N400 amplitude across the conditions (i.e., largest in SV and smallest in HC and medium in LC) in all of the three tasks (FONT,  $F(1, 26) = 11.22$ ,  $p < .01$ ; READ,  $F(1, 26) = 25.24$ ,  $p < .001$ ; SEM,  $F(1, 26) = 14.22$ ,  $p < .001$ ), indicating that there was a N400 modulation effect in all of the three tasks (Fig. 5A).

Further conventional statistical analysis showed a significant sentence type main effect ( $F(2, 52) = 14.94$ ,  $p < .001$ , with a larger N400 amplitude in the SV than in LC and HC, and in the LC than in the HC,  $ps < .05$ ), but no interaction between sentence type and task in the 300–500 ms time window. Although there was a significant interaction of the sentence type by task and by region ( $F(2, 52) = 2.88$ ,  $p < .05$ ), a larger N400 amplitude was shown in the SV than in the LC and HC in the posterior regions in all three tasks ( $ps < .05$ ), with an additionally larger N400 amplitude in LC than in the HC in READ ( $p < .05$ ).

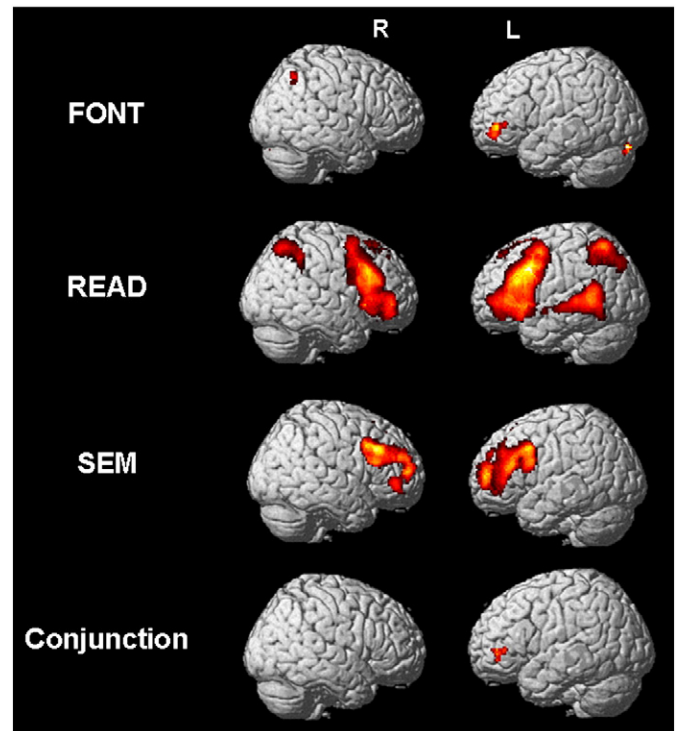
For the 500–800 and 800–1200 ms time windows, neither the sentence type main effect nor the interaction of sentence type by the task was shown. Nevertheless, a significant interaction of the sentence type by the region was found in both of the time windows, with a larger LPC amplitude in the SV than that in the LC ( $p < .06$ ) in the posterior region in the 500–800 ms time window, and larger LPC amplitude in the SV than in the LC ( $p < .05$ ) in the posterior region in the 800–1200 ms time window. More detailed information about conventional analysis is summarized in the appended Table S1.

#### Imaging results

As shown in Fig. 4, the parametric modulation analysis revealed a number of brain regions associated with the processing of the critical word. They include bilateral IFG, bilateral parietal cortex and left MTG for the reading task, bilateral IFG for the semantic task, and LIFG, right parietal cortex and visual cortex for the font size task (small volume correction, radius of sphere = 8 mm, corrected at 0.05 with FWE at cluster-level). Details of these activations are shown in Table 1. The conjunction analysis revealed a single region of activation in LIFG (small volume correction, radius of sphere = 8 mm, corrected at 0.05 with FWE at cluster-level). Note that the task by sentence type interaction in the fMRI data did not show any significant activation at the same threshold. We also checked the pairwise comparisons

results and found that the aLIFG activation revealed from the parametric modulation analysis was present in almost all of the pairwise contrasts in all of the tasks.

Fig. 5B shows the pattern of the fMRI BOLD signals in this region, which are separated by sentence type and task. To better capture the semantic unification load effect, we performed a linear trend



**Fig. 4.** The effect of parametric modulation on the BOLD signal associated with critical word processing. FONT for font size judgment task, READ for reading for comprehension task, SEM for semantic congruency judgment task, and Conjunction for the conjunction effect of FONT, READ and SEM. L for left and R for right.

**Table 1**  
Regions of activation modulated by semantic unification load in all three tasks.

| Task                    | Regions                   | BA    | MNI |     |     | Peak<br>t value | Voxel | FWE     |
|-------------------------|---------------------------|-------|-----|-----|-----|-----------------|-------|---------|
|                         |                           |       | x   | y   | z   |                 |       |         |
| READ                    | L SFG/Medial FG           | 6/8   | -2  | 20  | 54  | 9.37            | 1604  | 0.001   |
|                         | L IFG/MFG                 | 9/45  | -48 | 26  | 24  | 8.4             | 7090  | 0.001   |
|                         | R MFG/IFG                 | 46    | 50  | 20  | 30  | 7.24            | 5257  | 0.001   |
|                         | L MTG/STG                 | 21/22 | -56 | -48 | -4  | 6.13            | 2229  | 0.001   |
|                         | L SPL                     | 7     | -30 | -60 | 58  | 6.04            | 2070  | 0.001   |
|                         | R SPL/IPL/TPJ             | 7/40  | 32  | -66 | 56  | 4.8             | 975   | 0.003   |
| SEM                     | L SFG/MFG                 | 9/46  | -42 | 40  | 34  | 5.47            | 2719  | 0.001   |
|                         | L SFG/Bilateral Medial FG | 8/9   | -2  | 22  | 54  | 4.8             | 1260  | 0.001   |
|                         | R MFG                     | 10/46 | 32  | 56  | 12  | 4.71            | 2321  | 0.001   |
|                         | L IPL                     | 40    | -54 | -46 | 40  | 4.13            | 681   | 0.021   |
| FONT                    | L MFG/IFG                 | 10/45 | -42 | 48  | -2  | 3.34            | 204   | 0.014 * |
|                         | L Declive                 | /     | -18 | -84 | -26 | 3.58            | 151   | 0.026 * |
|                         | R SPL                     | 7     | 32  | -62 | 46  | 3.2             | 72    | 0.041 * |
|                         | R Posterior Lobe          | /     | 10  | -84 | -24 | 3.36            | 65    | 0.044 * |
| Three tasks conjunction | L IFG/MFG                 | 45/47 | -50 | 38  | 4   | 2.99            | 70    | 0.074 * |

Note: \*  $p < 0.05$ , FWE corrected at voxel-level for Small Volume Correction (VOI's sphere radius = 8 mm) at cluster-level. L, left; R, right; MFG: Middle Frontal Gyrus; SFG: Superior Frontal Gyrus; Medial FG: Medial Frontal Gyrus; IFG: Inferior Frontal Gyrus; STG: Superior Temporal Gyrus; MTG: Middle Temporal Gyrus; SPL: Superior Parietal Lobule; IPL: Inferior Parietal Lobule; TPJ: Temporal–Parietal Junction; IOG: Inferior Occipital Gyrus; MOG: Middle Occipital Gyrus; PHG: Parahippocampal Gyrus.

analysis. There was no significant interaction in the beta value in the confirmatory analysis. Moreover, there were linear trends across the conditions in all three tasks for a signal, as shown in Fig. 5B (FONT,  $F(1, 26) = 11.22, p < .01$ ; READ,  $F(1, 26) = 25.24, p < .001$ ; SEM,  $F(1, 26) = 14.22, p < .001$ ).

**Discussion**

To reveal the neural basis of semantic unification in this study, we constructed three types of sentences with different semantic unification loads. Three types of tasks that involved semantic unification were employed for a conjunction analysis to exclude the brain activations that would accompany, but be non-essential, to the semantic unification process.

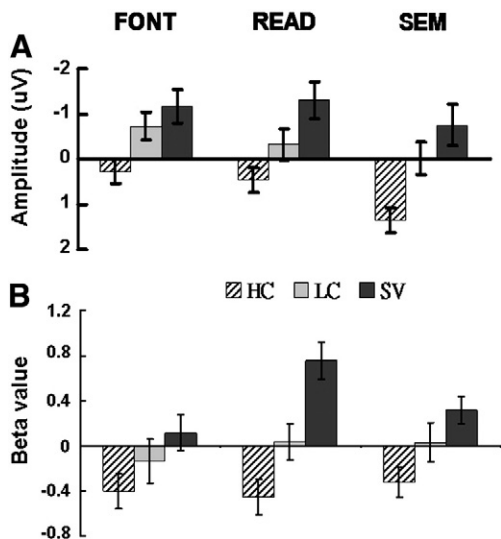
Consistent with previous findings, the incongruent sentences elicited a larger N400 than the congruent sentences, and the unexpected words in the sentences elicited a larger N400 amplitude than the expected words (Kutas and Federmeier, 2011). These N400 effects clearly indicated the presence of semantic unification in the SEM and READ tasks. Critically, the N400 effect was also observed for the FONT task. This finding, combined with the absence of the sentence

type by task interaction, suggests that semantic unification also occurred for the FONT task with comparable magnitude as in the other two tasks.

The fMRI results showed a clear modulation of the LIFG activation as the load of the semantic unification was varied across the sentence type in both the READ and SEM tasks. This was consistent with previous studies that identified the role of LIFG in semantic unification in explicit language tasks (Hagoort et al., 2009). Critically, the effect of sentence type was also found in the LIFG for the FONT task, although with a relatively weaker and less extensive activation. These results provide strong evidence that this anterior region of LIFG as identified in the conjunction analysis plays an important role in semantic unification. Thus, this activation is unlikely to be accounted for by other (e.g., task-related) processes that simply accompany semantic unification, as we mentioned in the Introduction. Such processes, being present in explicit tasks but not implicit tasks, would not produce a common region of activation. For example, this LIFG activation could not reflect general strategic processing as has been argued in some previous studies where the IFG damaged patients comprehended simple sentences but not complex sentences, which demanded a higher level of cognitive control (Van Petten and Luka, 2006). Note that there are other studies that failed to find a modulation of the LIFG activity by the level of difficulty of the semantic unification (Crinion et al., 2003; Vandenberghe et al., 2002), but the reason for these negative results is still unclear.

Note that the negative signal changes in the aLIFG in the HC condition in all of the three tasks should not be taken to indicate that the aLIFG was inactive during the semantic unification (see Ye and Zhou (2009) for a similar case of negative signals). First, to avoid potential confounding effects, such as the wrap-up effect at the end of the sentence, we modeled the critical word event with a short duration regressor (0.6 s), which may have underestimated the activation associated with the semantic unification. Second, we computed a BOLD signal value, which reflects the relative, but not absolute values, being that the positive or negative value depends on the baseline estimation (Gusnard and Raichle, 2001), as well as other predictors included in the statistical model.

One may use the LC minus the HC or a condition minus the baseline activation map as a mask to localize the semantic unification effect. However, the LC differs from the HC not only in semantic unification, but also in other processes. For example, the LC may trigger a higher general cognitive control to override the word expected from the context when the reader encountered the unexpected words, and such a control process may not be the same as semantic unification (cf. Hagoort et al., 2009). Thus, parametrically



**Fig. 5.** Amplitude in 300–500 ms time window in the posterior region (A) and beta value in the aLIFG (B) across the sentence types and tasks. HC for the high cloze sentence, LC for the low cloze sentence, and SV for the violation sentence.

manipulating the semantic unification load provides a more reasonable way to indicate involvement in the semantic unification effort.

Both the READ and SEM tasks required the participants to explicitly attend to word meanings, which may induce a controlled semantic retrieval and/or semantic selection that would not be present during the FONT task. This difference may explain the activation in the posterior LIFG for the two explicit tasks, but not the implicit task, as this region is known to be associated with controlled selection processes (Bilenko et al., 2009; Thompson-Schill et al., 2005; Zempleni et al., 2007). Such processes may not be language specific as a general cognitive control was also found to activate the posterior LIFG (January et al., 2009; Ye and Zhou, 2009).

The right IFG (RIFG) activation in the SEM and READ tasks had been previously reported (Vigneau et al., 2011; Zhu et al., 2009). One notion is that the RIFG has similar functional roles as the LIFG except that it is only engaged during a high working memory load (Vigneau et al., 2011). Consistent with this notion, we did find a semantic unification load that modulated the signal change in the RIFG in both the READ and SEM tasks.

The activation in the left S/MTG found in the READ task is consistent with the proposal that the posterior temporal areas are important for semantic processing (Hagoort, 2005; Jung-Beeman, 2005; Lau et al., 2008). An increased semantic unification load could enhance semantic retrieval, which in turn might engage more of the posterior temporal regions (Tesink et al., 2009). The other region activated in the READ task was the left inferior parietal lobe (IPL), a region that is often associated with semantic or phonological working memory (Gold et al., 2005), as well as general working memory load (Owen et al., 2005). Connectivity analysis revealed close coupling between the LIFG and IPL as well as between the IPL and the temporal cortex (Catani et al., 2005; Xiang et al., 2010). These results suggest that the posterior activation may be task-dependent. For example, during READ, the reader attempted to obtain a coherent representation of the sentences regardless of whether the sentence was congruent or not. Such efforts may include extra word-level semantic retrieval, engaging more working memory in the LC and SV than the HC, and consequently leading to the parametric modulation effect in READ. However, in FONT and SEM, such parametric modulation effects may be absent due to different reasons. In the Font task, there was actually no explicit requirement for constructing a coherent semantic representation. While in the SEM task, the participants may use a specific response strategy, terminating semantic processing (and related working memory processing) before the sentence ended as soon as they detected the incongruence (Andreatta et al., 2010). Consequently, the cognitive processes and the modulation effect in SEM were absent.

Contextual fit may affect processes other than semantic unification. In fact, it would be almost impossible to identify a single manipulation that would only affect the semantic unification but nothing else. The aim of this present study was to use the conjunction analysis across the three different types of tasks to exclude those other processes as much as possible. Nevertheless, positive component after the N400 time window was also observed in the 500–800 and 800–1200 ms windows. A late positive effect was often reported after the N400 when highly predicted words were replaced with other words (DeLong et al., 2011; review see Van Petten and Luka, 2012). A higher context constraint would elicit larger late positive components (Federmeier et al., 2007). It was also suggested that a high sentence constraint would predict a particular word, when the participants encountered the unexpected word they would need to suppress the expected one to comprehend the sentence (e.g., Federmeier et al., 2007). Generally, this positive effect may reflect the reprocessing cost (Van Petten and Luka, 2012). The late positive component was presented in the READ and SEM, but was absent in the FONT. Similarly, the posterior LIFG modulation was present in the READ and SEM, but was absent in the FONT, thereby suggesting that the two may be related. Although this is a

possibility, as the posterior LIFG is known to be sensitive to inhibition (Thompson-Schill et al., 2005; Wright et al., 2011), more research is needed to establish a direct link.

The parametric design was rarely used in the semantic unification literature, which usually contrasts the incongruent and congruent sentences, ambiguity and non-ambiguity or stronger ambiguity and weaker ambiguity sentences, to reveal brain regions for semantic unification. Combining a parametric design and a conjunction analysis, this study showed that the anterior part of the LIFG is indeed critical for semantic unification and that the activation in this region cannot be attributed to other cognitive processes that often accompany semantic unification.

Supplementary materials related to this article can be found online at doi:10.1016/j.neuroimage.2012.02.036.

## Acknowledgments

The work was supported by grants obtained from the Natural Science Foundation of China (30970894 and 31100811), the Chinese Ministry of Education Project of Key Research Institute of Humanities and Social Sciences in Universities (10JJDXXL005), as well as grants obtained from the Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University.

## References

- Andreatta, R.D., Stemple, J.C., Joshi, A., Jiang, Y., 2010. Task-related differences in temporo-parietal cortical activation during human phonatory behaviors. *Neurosci. Lett.* 484, 51–55.
- Bilenko, N.Y., Grindrod, C.M., Myers, E.B., Blumstein, S.E., 2009. Neural correlates of semantic competition during processing of ambiguous words. *J. Cogn. Neurosci.* 21, 960–975.
- Braze, D., Mencl, W.E., Tabor, W., Pugh, K.R., Constable, R.T., Fulbright, R.K., Magnuson, J.S., Van Dyke, J.A., Shankweiler, D.P., 2011. Unification of sentence processing via ear and eye: an fMRI study. *Cortex* 47, 416–431.
- Caplan, D., Chen, E., Waters, G., 2008. Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex* 44, 257–275.
- Catani, M., Jones, D.K., Fyfe, D.H., 2005. Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.
- Constable, R.T., Pugh, K.R., Berroya, E., Mencl, W.E., Westerveld, M., Ni, W., Shankweiler, D., 2004. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *NeuroImage* 22, 11–21.
- Crinion, J.T., Price, C.J., 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain* 128, 2858–2871.
- Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.J.S., 2003. Temporal lobe regions engaged during normal speech comprehension. *Brain* 126, 1193–1201.
- DeLong, K.A., Urbach, T.P., Groppe, D.M., Kutas, M., 2011. Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology* 48, 1203–1207.
- Devauchelle, A.D., Oppenheim, C., Rizzi, L., Dehaene, S., Pallier, C., 2009. Sentence syntax and content in the human temporal lobe: an fMRI adaptation study in auditory and visual modalities. *J. Cogn. Neurosci.* 21, 1000–1012.
- Diaz, M.T., Barrett, K.T., Hogstrom, L.J., 2011. The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia* 49, 320–330.
- Federmeier, K.D., Wlotko, E.W., De Ochoa-Dewald, E., Kutas, M., 2007. Multiple effects of sentential constraint on word processing. *Brain Res.* 1146, 75–84.
- Friederici, A.D., Ruschmeyer, S.A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb. Cortex* 13, 170–177.
- Friederici, A.D., Makuuchi, M., Bahlmann, J., 2009. The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport* 20, 563–568.
- Gold, B.T., Balota, D.A., Kirchoff, B.A., Buckner, R.L., 2005. Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from fMRI adaptation. *Cereb. Cortex* 15, 1438–1450.
- Gusnard, D.A., Raichle, M.E., 2001. Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694.
- Hagoort, P., 2005. On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9, 416–423.
- Hagoort, P., 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1055–1069.
- Hagoort, P., Brown, C., 1994. Brain responses to lexical ambiguity resolution and parsing. In: Clifton Jr., C., Frazier, L., Rayner, K. (Eds.), *Perspectives on Sentence Processing*. Lawrence Erlbaum Associates, Hillsdale NY, pp. 45–80.
- Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441.

- Hagoort, P., Baggio, G., Willems, R.M., 2009. Semantic unification. In: Gazzaniga, M.S. (Ed.), *The Cognitive Neurosciences*. MIT Press, London, pp. 819–836.
- Huang, J., Zhu, Z., Zhang, J.X., Wu, M., Chen, H.-C., Wang, S., 2012. The role of left inferior frontal gyrus in explicit and implicit semantic processing. *Brain Res.* 1440, 56–64.
- Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2007. Time course of semantic processes during sentence comprehension: an fMRI study. *NeuroImage* 36, 924–932.
- January, D., Trueswell, J.C., Thompson-Schill, S.L., 2009. Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *J. Cogn. Neurosci.* 21, 2434–2444.
- Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. *Trends Cogn. Sci.* 9, 512–518.
- Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. *Science* 274, 114–116.
- Kaan, E., Swaab, T.Y., 2003. Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110.
- Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49.
- Kuperberg, G.R., Holcomb, P.J., Sitnikova, T., Greve, D., Dale, A.M., Caplan, D., 2003. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J. Cogn. Neurosci.* 15, 272–293.
- Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann. Rev. Psychol.* 62, 621–647.
- Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.
- Mason, R.A., Just, M.A., 2007. Lexical ambiguity in sentence comprehension. *Brain Res.* 1146, 115–127.
- Newman, S.D., Lee, D., Ratliff, K.L., 2009. Off-line sentence processing: what is involved in answering a comprehension probe? *Hum. Brain Mapp.* 30, 2499–2511.
- Ni, W., Constable, R.T., Mencl, W.E., Pugh, K.R., Fulbright, R.K., Shaywitz, S.E., Shaywitz, B.A., Gore, J.C., Shankweiler, D., 2000. An event-related neuroimaging study distinguishing form and content in sentence processing. *J. Cogn. Neurosci.* 12, 120–133.
- Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59.
- Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88.
- Relander, K., Rama, P., Kujala, T., 2009. Word semantics is processed even without attentional effort. *J. Cogn. Neurosci.* 21, 1511–1522.
- Rodd, J.M., Davis, M.H., Johnsrude, I.S., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb. Cortex* 15, 1261–1269.
- Rogalsky, C., Hickok, G., 2009. Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb. Cortex* 19, 786–796.
- Rueschemeyer, S.A., Zysset, S., Friederici, A.D., 2006. Native and non-native reading of sentences: an fMRI experiment. *NeuroImage* 31, 354–365.
- Tesink, C.M., Petersson, K.M., van Berkum, J.J., van den Brink, D., Buitelaar, J.K., Hagoort, P., 2009. Unification of speaker and meaning in language comprehension: an fMRI study. *J. Cogn. Neurosci.* 21, 2085–2099.
- Thompson-Schill, S.L., Bedny, M., Goldberg, R.F., 2005. The frontal lobes and the regulation of mental activity. *Curr. Opin. Neurobiol.* 15, 219–224.
- Van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.* 97, 279–293.
- Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: Benefits, costs, and ERP components. *Int J Psychophysiol.* 83, 176–190.
- Vandenberghe, R., Nobre, A.C., Price, C.J., 2002. The response of left temporal cortex to sentences. *J. Cogn. Neurosci.* 14, 550–560.
- Vigneau, M., Beaucousin, V., Herve, P.Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2011. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage* 54, 577–593.
- Wang, S., Li, N., Yang, J., 2011. Semantic preview effect and high level semantic processing in Chinese sentence comprehension. The 7th Chinese Psychologist Conference, September, Taipei.
- Warren, J.E., Crinion, J.T., Lambon Ralph, M.A., Wise, R.J., 2009. Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. *Brain* 132, 3428–3442.
- Wright, P., Randall, B., Marslen-Wilson, W.D., Tyler, L.K., 2011. Dissociating linguistic and task-related activity in the left inferior frontal gyrus. *J. Cogn. Neurosci.* 23, 404–413.
- Xiang, H.D., Fonteijn, H.M., Norris, D.G., Hagoort, P., 2010. Topographical functional connectivity pattern in the perisylvian language networks. *Cereb. Cortex* 20, 549–560.
- Ye, Z., Zhou, X., 2009. Conflict control during sentence comprehension: fMRI evidence. *NeuroImage* 48, 280–290.
- Zempleni, M.Z., Renken, R., Hoeks, J.C.J., Hoogduin, J.M., Stowe, L.A., 2007. Semantic ambiguity processing in sentence context: evidence from event-related fMRI. *NeuroImage* 34, 1270–1279.
- Zhu, Z., Zhang, J.X., Wang, S., Xiao, Z., Huang, J., Chen, H.-C., 2009. Involvement of left inferior frontal gyrus in sentence-level semantic integration. *NeuroImage* 47, 756–763.
- Zhu, Z., Wang, S., Feng, G., Liu, Y., 2011. Left inferior frontal gyrus and semantic unification in sentence comprehension. *Adv. Psychol. Sci.* 19, 1147–1157.