

Unification of Speaker and Meaning in Language Comprehension: An fMRI Study

Cathelijne M. J. Y. Tesink^{1,2}, Karl Magnus Petersson^{1,3},
Jos J. A. Van Berkum^{1,3,4}, Daniëlle van den Brink^{1,2},
Jan K. Buitelaar^{2,5}, and Peter Hagoort^{1,3}

Abstract

■ When interpreting a message, a listener takes into account several sources of linguistic and extralinguistic information. Here we focused on one particular form of extralinguistic information, certain speaker characteristics as conveyed by the voice. Using functional magnetic resonance imaging, we examined the neural structures involved in the unification of sentence meaning and voice-based inferences about the speaker's age, sex, or social background. We found enhanced activation in the inferior frontal gyrus bilaterally (BA 45/47) during listening to sentences whose meaning was incongruent with inferred

speaker characteristics. Furthermore, our results showed an overlap in brain regions involved in unification of speaker-related information and those used for the unification of semantic and world knowledge information [inferior frontal gyrus bilaterally (BA 45/47) and left middle temporal gyrus (BA 21)]. These findings provide evidence for a shared neural unification system for linguistic and extralinguistic sources of information and extend the existing knowledge about the role of inferior frontal cortex as a crucial component for unification during language comprehension. ■

INTRODUCTION

During speech comprehension, the human brain derives an interpretation of the speaker's message by integrating different sources of information. In psycholinguistic models, phonology, syntax, and semantics are seen as the core aspects of our language faculty, and the extraction of meaning from speech requires continuous and parallel use of information related to these linguistic information sources. In addition, for the listener's understanding of a speaker's message, it is essential that the brain unifies sentence meaning with other sources of information that contribute to the understanding of a spoken utterance. In the current functional magnetic resonance imaging (fMRI) study, we focused on the unification of sentence meaning with one particular source of extralinguistic information that is inherent to speech, information about speaker characteristics conveyed by the voice.

When listening to an unknown and invisible speaker, for instance on the telephone, one not only hears the content of the speaker's message but also derives information about this speaker from the voice, such as her sex, age, and social class. Thus, not only is the human voice

the carrier of linguistically coded information, it also implicitly conveys important nonlinguistic information concerning speaker characteristics. Functional neuroimaging studies have revealed cortical regions that are selectively sensitive to the human voice (see, for a review, Belin, Fecteau, & Bedard, 2004). These voice-sensitive regions can be found bilaterally along the upper bank of the superior temporal sulcus (STS) and they appear to respond significantly more to human vocal sounds, whether speech or nonspeech, than to other naturally occurring sounds, such as nonhuman and nonvocal sounds (Fecteau, Armony, Joanette, & Belin, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000).

Although both anterior STS regions are sensitive to vocal sounds, each has a slightly different contribution. Left, but not right, anterior STS regions display stronger activation when (intelligible) linguistic information is present in the voice than when the auditory input consists of nonspeech vocalizations (Belin, Zatorre, & Ahad, 2002; Scott, Blank, Rosen, & Wise, 2000). Conversely, the right anterior STS shows a stronger voice-sensitive response to nonspeech vocal sounds such as laughs and cries, and does not seem to require verbal content to be responsive to vocal sounds (von Kriegstein & Giraud, 2004; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; Belin et al., 2002). In addition, the right anterior STS is involved in processing speaker identity characteristics in the voice, or more specifically, in speaker recognition

¹Radboud University Nijmegen, Nijmegen, ²Radboud University Nijmegen Medical Centre, Nijmegen, ³Max Planck Institute for Psycholinguistics, Nijmegen, ⁴University of Amsterdam, Amsterdam, ⁵Karakter Child and Adolescent Psychiatry University Centre, Nijmegen

(von Kriegstein & Giraud, 2004; Belin & Zatorre, 2003; von Kriegstein et al., 2003).

Although research has been done on voice perception, the issue of which brain regions support the unification of speaker characteristics inferred from the voice with semantic information in speech has not been addressed. Recently, a model that deals with the unification of different aspects of language related information in the brain was put forward (Hagoort, 2005). This framework distinguishes three functional components of language processing: memory (mental lexicon), unification (integration), and control. In the context of our study, the unification component is the most relevant of the three. Unification refers to the on-line integration of lexical information that is retrieved from memory (i.e., from the mental lexicon) into a representation of a multiword utterance. It is suggested that during language comprehension, as well as production, unification operations take place in parallel and interactively at the semantic, syntactic, and phonological levels of language processing (Jackendoff, 2007). Furthermore, the abovementioned model argues that the left inferior frontal gyrus (LIFG) is a crucial brain region for unification (Hagoort, 2005).

A number of neuroimaging studies have found evidence for the role of left inferior frontal cortex in semantic unification, that is, in the integration of word meaning into an unfolding representation of the sentence context. Studies investigating semantic unification are often based on the rationale that sentences containing semantic anomalies or ambiguous words have a higher semantic unification load than correct sentences because in anomalous sentences more effort is needed to integrate word information into the sentence context (Rodd, Davis, & Johnsrude, 2005; Hagoort, Hald, Bastiaansen, & Petersson, 2004). In functional neuroimaging studies exploiting this paradigm, an increased BOLD response in the LIFG (Brodmann's area's [BA] 45/47) was observed for sentences containing a semantic anomaly (Hagoort et al., 2004; Ni et al., 2000) or ambiguity (Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007; Rodd et al., 2005). Furthermore, manipulation of the semantic unification load by presenting sentences with semantic or world knowledge anomalies has revealed that the semantic unification area in the LIFG is not only involved in determining whether an interpretation is semantically coherent but it is also recruited to verify the meaning of an utterance in relation to our knowledge of the world (Hagoort et al., 2004). Importantly, increased activation in the LIFG has also been observed for correct sentences, that is, sentences without any anomaly, relative to a low-level baseline (Hagoort, 2005; Ni et al., 2000). Taken together, these findings suggest that the LIFG is recruited during semantic unification. We would like to stress that there does not exist a true dichotomy between "correct" sentences and sentences with "anomalies." Rather, there is a continuum from sentences that fit very well with our knowledge about the world and about our language to

sentences that are not compatible with what we know. Ultimately, language allows us to communicate not only what we already know but also what we did not know (i.e., new information). However, for the sake of simplicity, the terms "correct" and "anomaly" will be used.

So far we have used the term "unification" to refer to the on-line assembly of complex meaning during language comprehension. Although the term "integration" is often used as a synonym for unification, we suggest that it is useful to make a functional distinction between the two (Hagoort, Baggio, & Willems, in press). Semantic *integration* occurs if different sources of information converge on a common memory representation, for example, the sound and the sight of an animal (e.g., a meowing cat). The sight of a cat, the meowing sound, and their combined occurrence most likely all activate a memory representation of "cat" that is multimodal in nature. Semantic *unification*, on the other hand, is always a constructive process in which a semantic representation is built up that is not already stored in memory. Importantly, this distinction makes opposite predictions for the BOLD response. Semantic unification is always harder for semantic incongruities. The increased unification load for semantic incongruities should result in a stronger BOLD response than when semantically congruent items are presented. In contrast, during integration, congruent input provides converging support for a prestored representation, which might then be more strongly activated compared to a situation with incongruent input (Hagoort et al., in press). Hence, in the case of integration, the congruent condition will elicit a stronger BOLD response than the incongruent condition. A few studies on multimodal integration have indeed reported activation increases in superior temporal cortex to matching stimulus combinations (van Atteveldt, Formisano, Blomert, & Goebel, 2007; Calvert, Campbell, & Brammer, 2000).

In this fMRI study, we investigated the neural underpinnings of unifying the meaning of a spoken sentence with extralinguistic information conveyed by the speaker's voice. We presented participants with spoken sentences whose meaning did (speaker-congruent) or did not (speaker-incongruent) match inferences of the listener about the speaker's age, sex, or social background that were based on the speaker's voice. In the speaker-incongruent sentences, there was one specific word at which the sentence became harder to interpret given the speaker's characteristics as inferred from the voice (printed in italics in the examples that follow). Examples: "Every evening I drink a glass of *wine* before going to bed" in a young child's voice, "My favorite colors are *pink* and lime green" in a male voice, and "I have a large *tattoo* on my back" spoken in an upper-class accent. By manipulating the congruency of sentence meaning and voice-based inferences about a speaker, we were able to identify brain regions that are responsive to variations in unification load. We also included sentences with stan-

dard semantic or world knowledge anomalies to examine whether there are common brain regions involved in unifying linguistic and extralinguistic information.

An experiment with event-related brain potentials (ERPs) and the same materials as used in this study found that ERPs time-locked to the critical words elicited an N400 effect in the speaker-incongruent condition (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). The N400 effect is an amplitude modulation of the N400 component that is sensitive to semantic anomalies as well as to subtle manipulations in semantic integration processes (Hagoort & Brown, 1994; Kutas & Hillyard, 1980, 1984). In addition, the results revealed that speaker-incongruent sentences elicited the same type of N400 effect as semantic or world knowledge anomalies. This suggests that voice-based inferences about the speaker affect the same early interpretation mechanism that is sensitive to lexical-semantic and world knowledge information.

Given the ERP and fMRI results reviewed above, we predict that unification of sentence meaning and extralinguistic information from the speaker's voice will engage the same brain region in the LIFG as the unification of lexical-semantic information and world knowledge. Although we mainly expect that the LIFG is recruited during unification, it is not uncommon to find homologue regions to be activated during language tasks. In line with what has been found when unification is studied in a discourse context (Menenti, Petersson, Scheeringa, & Hagoort, in press), it is well possible that we will observe activation in the right inferior frontal gyrus (RIFG) during unification operations. Our findings extend the existing knowledge about the role of left inferior frontal cortex as a crucial component for unification of linguistic and nonlinguistic information, such as, for example, for social information.

METHODS

Participants

Forty-three healthy right-handed native speakers of Dutch participated in the experiment of whom 42 were included in the final analysis (18 women; mean age \pm $SD = 23.9 \pm 4.6$ years). All participants had normal or corrected-to-normal vision and normal hearing. None of them used any medication, had a history of head trauma, or neurological or psychiatric illness. Written informed consent was obtained according to the Declaration of Helsinki. One of the participants was excluded from the final analysis because of excessive head movement.

Stimulus Material

The stimulus materials consisted of two sets of sentences: a set of speaker-inference sentences and a set of sentences with semantic or world knowledge anomalies (see Hagoort et al., 2004). The stimulus materials used in this

study were identical to those of the ERP study by Van Berkum et al. (2008).

For the set of speaker-inference sentences, we constructed 160 sentences with a lexical content that was congruent with voice-based inferences about a particular speaker, but incongruent with inferences about another speaker. To increase variability and to cover a broad range of speaker information captured in the voice, sentence meaning could be incongruent with respect to three different speaker characteristics: age, sex, or social background. In total, there were six types of speaker-incongruent utterances: 40 sentences were odd when pronounced by a male speaker ("My favorite colors are *pink* and lime green"), 40 sentences were odd when pronounced by a female speaker ("On Saturdays I work as a *bouncer* in a club"), 20 sentences were odd when pronounced by a child ("Every evening I drink a glass of *wine* before going to bed"), 20 sentences were odd when pronounced by an adult ("I cannot sleep without my *teddy bear* in my arms"), 20 sentences were odd when pronounced by a speaker with a Dutch accent that is associated with an upper-class background ("I have a large *tattoo* on my back"), and 20 were odd when pronounced by a speaker with Dutch accent that is associated with a lower-class background ("In my free time I enjoy listening to *piano music* by Chopin"). The sentences were created in such a way that the speaker incongruity always emerged at a specific word in the sentence, the critical word (here in italics), which was never sentence-final. Although some incongruities between voice-based inferences about the speaker's characteristics and sentence content were truly anomalous, the majority merely violated (Dutch) social stereotypes. Furthermore, the fragment before the critical word was compatible with either speaker ("Yesterday I went to...", "I have a large...").

We recorded the speaker-inference sentences with a total of 16 speakers (4 men and 4 women, 2 children aged 6 and 8 years and 2 adults, 2 speakers with a Dutch accent typically perceived as lower-class, and 2 with a Dutch accent typically perceived as upper-class). We selected recordings in which the congruent and incongruent variant of an item were pronounced with a similar prosodic contour. Furthermore, we matched speaker-congruent and speaker-incongruent recordings on: (1) acoustic duration of the critical words (speaker-congruent: mean = 520 msec, standard deviation [SD] = 149 msec, range = 236–1023 msec; speaker-incongruent: mean = 524 msec, SD = 140 msec, range = 212–921 msec); (2) duration of the preceding sentence fragment (speaker-congruent: mean = 1596 msec, SD = 492 msec, range = 485–3367 msec; speaker-incongruent: mean = 1629 msec, SD = 507 msec, range = 455–3261 msec); (3) sentence length (speaker-congruent: mean = 3182 msec, SD = 614 msec, range = 1638–5648 msec; speaker-incongruent: mean = 3228 msec, SD = 629 msec, range = 1784–5509 msec).

To investigate whether overlapping brain regions are involved in the unification of speaker information and unification of semantic information and world knowledge, we included an additional set of 36 triplets of sentences. Within a triplet, the sentences were identical with the exception of one critical word. Each triplet comprised a sentence that was semantically coherent (correct condition: “Dutch trains are *yellow* and blue”), a sentence that contained a semantic anomaly (“Dutch trains are *sour* and blue”), and a sentence with a world knowledge anomaly (“Dutch trains are *white* and blue”; see Hagoort et al., 2004 for details). The world knowledge sentences were recorded with four female speakers and one male speaker. The three items of a sentence triplet (“Dutch trains are *yellow/sour/white* and blue”) were always pronounced by the same speaker and their critical words were matched across conditions on: (1) acoustic duration (correct: mean = 431 msec, $SD = 109$ msec; semantic anomaly: mean = 425 msec, $SD = 94$ msec; world knowledge anomaly: mean = 451 msec, $SD = 133$ msec); (2) word frequency (on 3.7 million, Corpus Spoken Dutch R6) (correct: mean = 136 msec, $SD = 206$ msec; semantic anomaly: mean = 115 msec, $SD = 190$ msec; world knowledge anomaly: mean = 121 msec, $SD = 200$ msec); (3) duration of the preceding sentence fragment (correct: mean = 1870 msec, $SD = 517$ msec; semantic anomaly: mean = 1842 msec, $SD = 507$ msec; world knowledge anomaly: mean = 1841 msec, $SD = 508$ msec); and sentence length (correct: mean = 3302 msec, $SD = 656$ msec; semantic anomaly: mean = 3277 msec, $SD = 642$ msec; world knowledge anomaly: mean = 3303 msec, $SD = 649$ msec).

Forty-two items consisting of reversed speech were inserted as filler sentences. These items were included for a study on language processing in adults with an autism spectrum disorder and will not be analyzed for the research question of the present study.

Overall, the experimental sentences varied in length from 1638 to 5648 msec, with the average sentence length being 3247 msec ($SD = 597$). The critical words had an average duration of 480 msec ($SD = 136$ msec).

We created six different pseudorandomized trial lists such that each list contained an equal number of items per condition (80 speaker-incongruent and 80 speaker-congruent sentences, 36 sentences with a semantic anomaly, 36 sentences with a world knowledge anomaly, 36 correct sentences, and 42 reversed speech items). Furthermore, the items were distributed such that none of the participants heard more than one variant of the same sentence, with the constraint that no more than two items of the same condition were presented consecutively, and such that each speaker pronounced an equal number of congruent and incongruent sentences (for the speaker-inference sentences five of each type per speaker).

The materials of the present experiment were validated in a posttest in which an independent group of participants (12 men and 12 women) listened to the six stimu-

lus lists and were asked to rate on a 5-point scale “how normal or strange you think it is to have the speaker say this particular thing” (1 = completely normal, 5 = very strange; see also Van Berkum et al., 2008). As expected, utterances that contained a speaker incongruity were rated as less plausible (mean = 3.5, $SD = 0.8$, range = 1.5–5.0), than the corresponding speaker-congruent sentences (mean = 1.6, $SD = 0.4$, range = 1.0–3.1). Furthermore, utterances containing semantic anomalies were rated as highly implausible (mean = 4.6, $SD = 0.3$, range = 3.6–5.0), sentences with world knowledge anomalies were also rated as very implausible (mean = 4.2, $SD = 0.5$, range = 2.9–5.0), whereas the corresponding control sentences were perceived as acceptable (mean = 1.5, $SD = 0.4$, range = 1.0–2.6). The average semantic and world knowledge anomaly were considered to be more anomalous than the average speaker incongruity.

Experimental Design and Procedure

Each participant listened to a total of 314 sentences that were presented in an event-related design. During image acquisition, subjects lay in a supine position in the MR scanner and head movements were minimized by an adjustable padded head holder. The spoken sentences were presented through nonmagnetic headphones (Commander XG; Resonance Technology, Northridge, CA; www.mrvideo.com), which dampened scanner noise. The fixation cross was presented via an LCD projector standing outside the scanner room, projecting the computer display onto a semitransparent screen that the subject viewed through a mirror device attached to the head coil. Stimulus presentation was controlled by a PC running the Presentation software (version 9.70; Neurobehavioral Systems, San Francisco, CA; nbs.neuro-bs.com). Participants were instructed to process each sentence attentively for comprehension. To ensure attentive listening, they were told that afterward questions would be asked about the presented sentences. Before the beginning of the experiment, each participant received a practice block consisting of 10 sentences. These items were also used to adjust the volume level for sentence presentation. The scanner was switched on during the practice run and participants were asked to indicate whether the volume should go up or down. The volume level that suited each participant best was used in the experiment. The functional data acquired during the practice run were not used in the analysis.

Each trial began with a fixation asterisk presented in the center of the screen. After 300 msec the fixation asterisk disappeared for 1000 msec and then returned to indicate that the sentence was about to start. During sentence presentation, the asterisk remained on the screen and it lasted after sentence onset until the end of the trial. Trial onset was effectively jittered by adding 0, 500,

1000, or 1500 msec (mean = 750 msec) to the standard trial duration of 8200 msec. The experiment was divided into two blocks of 157 sentences each. Following the first block of sentences, there was a short break. At the start of each experimental block, we inserted two filler items (neutral sentences) to minimize loss of data due to saturation transients at the beginning of each block.

MRI Data Acquisition

During the listening task, we acquired whole head T2*-weighted EPI-BOLD fMRI data with a SIEMENS 1.5-T MR scanner using an ascending slice acquisition sequence (volume TR = 2440 msec, TE = 40 msec, 90° flip angle, 31 axial slices, slice-matrix size = 64 × 64, slice thickness = 3 mm, slice gap = 0.5 mm, field of view = 224 mm, isotropic voxel size = 3.5 × 3.5 × 3.5 mm³). Following the experimental session, a high-resolution structural MR image was acquired for each participant, using a T1-weighted MP-RAGE sequence (volume TR = 2250 msec, TE = 3.93 msec, 15° flip angle, 176 sagittal slices, slice-matrix size = 256 × 256, slice thickness = 1 mm, no slice gap, field of view = 256 mm).

MRI Data Analysis

Image preprocessing and statistical analysis were performed using SPM2 (www.fil.ion.ucl.ac.uk/spm/software/spm2). The first five volumes of each participant's dataset were discarded to allow for T1 equilibration. The functional EPI-BOLD images were realigned, and the subject-mean functional MR images were coregistered with the corresponding structural MR images. These images were subsequently slice-time corrected, spatially normalized (i.e., the normalized transformations were generated from the structural MR images and applied to the functional MR images), and transformed into a common space, as defined by the SPM Montreal Neurological Institute (MNI) T1 template. The functional EPI-BOLD images were then spatially filtered by convolving the functional images with an isotropic 3-D Gaussian kernel (10 mm FWHM).

The fMRI data were then statistically analyzed using the general linear model and statistical parametric mapping (Friston et al., 1995). At the first level, single-subject fixed effect analyses were conducted. Two models were tested in each participant's data separately: one with the experimental conditions speaker-congruent and speaker-incongruent and a second model included the three world knowledge conditions (sentences with a semantic anomaly, sentences with a world knowledge anomaly, correct sentences). These linear models included regressors to model the duration of the sentence presentation from the onset of the critical word to the end of the trial. We then temporally convolved the explanatory variables

with the canonical hemodynamic response function provided by SPM2. To remove any signal changes due to head motion, we included six realignment parameters describing the head movements as confounds in the model. The data were high-pass filtered to account for various low-frequency effects. Temporal autocorrelation was modeled as a first-order autoregressive AR(1)+ noise process. For the second-level analysis, the generated single-subject contrast images for the main effects were entered in a random effects analysis.

Region-of-interest Analyses

Given our a priori hypothesis regarding the role of the LIFG as the primary focus of interest, a region-of-interest analysis (ROI) was performed. A meta-analysis (Bookheimer, 2002) has shown that semantic processing is centered at the coordinates [−42, 25, 4] (Talairach & Tournoux, 1988), with a mean distance to the local maxima to this center coordinate of 15 mm (Pettersson, Forkstam, & Ingvar, 2004). Accordingly, we converted these Talairach coordinates to MNI coordinates and applied small-volume correction using a spherical ROI with a radius of 15 mm around [−42, 26, 6], thresholded at $p = .001$ (uncorrected).

Whole-brain Analysis

In addition to testing condition effects in the ROI, we also tested for the presence of other regions that were differentially activated by the experimental conditions. In the explorative whole-brain search, the results of the random effects analyses were thresholded at $p < .001$ (uncorrected). We employed cluster size as the test-statistic for our whole-brain analyses and only considered activation clusters significant at a threshold of $p < .05$ (corrected for multiple nonindependent comparisons). All local maxima are reported as MNI coordinates. Relevant anatomical landmarks and Brodmann's areas were identified using the atlas of the human brain (Mai, Assheuer, & Paxinos, 2004), the Anatomy Toolbox (Eickhoff et al., 2005; Amunts, Malikovic, Mohlberg, Schormann, & Zilles, 2000) and the Talairach Daemon (Lancaster et al., 2000).

RESULTS

Because the main focus of this study is on the unification of speaker characteristics inferred from the voice with sentence content, we will first report results related to this experimental manipulation. Then, we will go into effects found for the semantic and world knowledge conditions and look into brain regions that are common for the unification of semantic knowledge, world knowledge, and speaker characteristics.

Speaker-inference Sentences

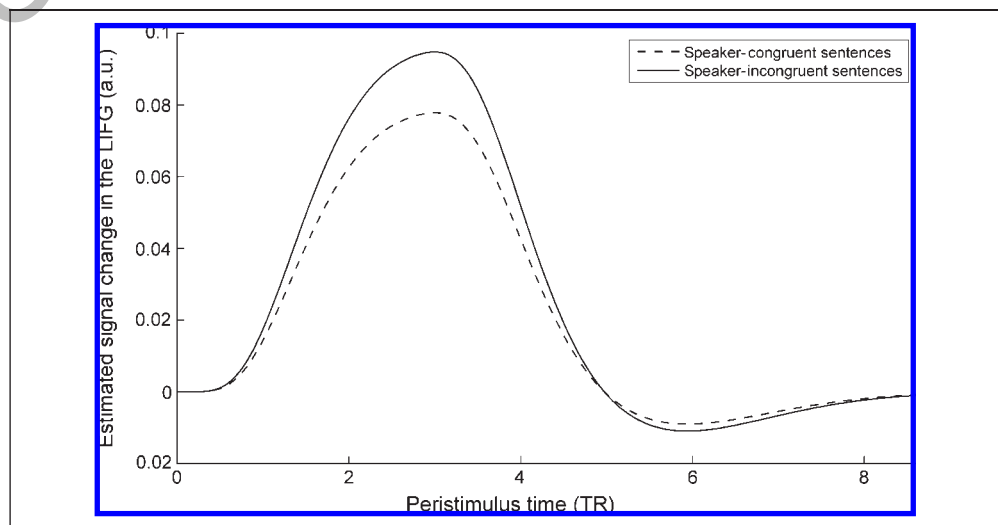
Region-of-interest Analysis

We first investigated whether the LIFG (BA 45/47) responded differently to sentences in which speaker characteristics conveyed by the voice were incongruent with sentence content. Using the ROI described in the Methods section, we found that the LIFG (BA 45/47) was activated significantly more strongly during speaker-incongruent sentences compared to speaker-congruent sentences [$t(41) = 4.5, p = .001$]. This effect corresponds to the predicted activation pattern and supports the hypothesis that the LIFG plays a role in unification of speaker characteristics during auditory sentence comprehension. Extracting BOLD responses for speaker-incongruent and speaker-congruent sentences in the LIFG showed that both conditions elicited an increase in activation in this region. This result is displayed in Figure 1 and is in line with earlier findings that showed that the LIFG is not only recruited during the processing of sentences with anomalies but is also implicated in the comprehension of coherent sentences (Willems, Ozyurek, & Hagoort, 2007; Hagoort et al., 2004).

Whole-brain Analysis

In the whole-brain analysis, a comparison was made between speaker-congruent and speaker-incongruent sentences. The results of the two contrasts speaker-incongruent > speaker-congruent and speaker-congruent > speaker-incongruent are listed in Table 1a and b. Next to significant activation in the LIFG, speaker-incongruent sentences elicited significant activation in the RIFG (BA 47; see Table 1a). A region in the posterior part of the left middle temporal gyrus (BA 21) showed a trend to respond more strongly to speaker-incongruent sentences than to speaker-congruent sentences ($p = .077$). Figure 2A displays renderings with the clusters of activation for the speaker-incongruent sentences.

Figure 1. Fitted BOLD responses for speaker-incongruent and speaker-congruent sentences from the ROI in the LIFG (15 mm sphere, center $[-42, 26, 6]$). This figure shows the involvement of the LIFG in the processing of both speaker-incongruent and speaker-congruent sentences.



Inspection of regions that were activated significantly stronger for speaker-congruent compared to speaker-incongruent sentences revealed activation clusters in bilateral superior temporal cortex (BA 22) extending into the anterior transverse temporal (Heschl's) gyrus (BA 41), in the right lingual gyrus (BA 18), and in right posterior cingulate cortex [PCC] (BA 29; see Table 1b for a complete list). Figure 2B shows renderings displaying the activation for speaker-congruent sentences.

World Knowledge and Semantic Anomalies

Region-of-interest Analysis

A previous study on unification of world knowledge and semantic information showed that unification of both sorts of information involved the LIFG (BA 45/47; Hagoort et al., 2004). To inspect whether our results were in line with these findings, we performed a small-volume correction using the ROI that was also used for the speaker-inference sentences. For the world knowledge sentences, we found that the LIFG (BA 45/47) was activated significantly more strongly during listening to sentences with a world knowledge anomaly compared to correct sentences [$t(82) = 4.93, p = .001$]. Also for the semantic contrast, the LIFG (BA 45/47) was significantly more activated during sentences containing a semantic anomaly than during correct sentences [$t(82) = 7.40, p = .001$]. Thus, the effects found in the ROI analyses for sentences with semantic and world knowledge anomalies replicate earlier findings of the study by Hagoort et al. (2004), in which sentences were presented visually.

Whole-brain Analysis

In the whole-brain analysis, a comparison was made between sentences with a world knowledge anomaly and correct sentences. The contrast world knowledge anomaly > correct sentences revealed significantly stronger

Table 1.

<i>Anatomical Region</i>	<i>BA</i>	<i>Cluster Size</i>	<i>Voxel T Value</i>	<i>MNI Coordinates</i>		
				<i>x</i>	<i>y</i>	<i>z</i>
<i>a. Results from the Whole-brain Analysis for the Contrast Speaker-incongruent > Speaker-congruent Sentences*</i>						
LIFG (pars triangularis)	45	398	4.50	-54	26	14
	45/47		4.22	-48	26	-2
	45		4.07	-50	22	8
RIFG (pars orbitalis)	47	211	4.48	50	34	-12
	47		4.47	48	24	-14
	47		4.30	54	28	-6
L. Middle temporal gyrus**	21	150	4.23	-62	-36	-8
	21		4.18	-58	-42	-4
<i>b. Results from the Whole-brain Analysis for the Contrast Speaker-congruent > Speaker-incongruent Sentences*</i>						
R. Anterior transverse temporal gyrus	41	515	5.49	38	-28	12
	41		4.62	46	-24	6
R. STG	22		4.04	58	-10	6
L. Anterior transverse temporal gyrus	41	706	4.94	-44	-26	6
L. Planum temporale	42		4.53	-60	-16	12
L. STG	22		4.33	-52	-10	4
R. Lingual gyrus	18	211	4.72	10	-54	4
R. PCC	29		3.50	6	-44	10

*Table shows all clusters at a significance level of $p < .05$ corrected at cluster-level (first thresholded at $p < .001$, uncorrected). All local maxima are reported as MNI coordinates. Significant activation peaks > 8 mm apart.

** $p = .077$, corrected at cluster level.

activation for sentences with a world knowledge anomaly in the LIFG (BA 45/47), left middle frontal gyrus (BA 6/9), left inferior and middle temporal gyrus (BA 20/21/22; see Table 2a for the complete list). Furthermore, in the right hemisphere, there was a significant cluster of activation in the middle temporal gyrus (BA 21/22) extending into the superior temporal gyrus (STG; BA 38). The reversed contrast, correct sentences $>$ world knowledge anomaly, showed significantly increased activation for correct sentences in right middle and left PCC (BA 24/31; see Table 2b). Figure 3 displays renderings showing activation for sentences with world knowledge anomalies.

Moreover, relative to correct sentences, sentences with a semantic anomaly elicited significantly increased activation in the LIFG (BA 45/47), left middle frontal gyrus (BA 6/9), left middle temporal gyrus (BA 21/22), and STG bilaterally (BA 21/22; see Table 3a). The contrast correct sentences $>$ sentences with a semantic anomaly showed significant clusters of activation in the inferior rostral gyrus bilaterally (BA 10/11/12), right posterior cingulate gyrus (BA 31), and angular gyrus bilaterally (BA 39/19;

Table 3b). Figure 4 shows renderings with the activation clusters for sentences with semantic anomalies.

Common Brain Regions for Unification of Speaker Characteristics, Semantic Knowledge, and World Knowledge

To test for common neural correlates for the unification of world knowledge, semantic knowledge, and speaker characteristics, we created a contrast showing regions involved in semantic and world knowledge anomalies relative to correct sentences. The resulting image of this contrast was then used to perform a small-volume correction on the contrast speaker-incongruent $>$ speaker-congruent sentences. The results of this analysis are displayed in Table 4. Brain regions involved in unification of world knowledge, semantic information, and speaker characteristics were the LIFG (BA 45/47), the RIFG (BA 47), and the posterior part of the left middle temporal gyrus (BA 21). These findings confirm the idea that there is an overlap in brain regions involved in the unification of linguistic and extralinguistic information.

Figure 2. (A) Speaker-incongruency effect. Activation clusters from the whole-brain analysis for the speaker-incongruent sentences relative to the speaker-congruent sentences, pooled across speaker dimensions (age, sex, social background). (B) Speaker-congruency effect. Activation clusters from the whole-brain analysis for the speaker-congruent sentences relative to the speaker-incongruent sentences, pooled across speaker dimensions.

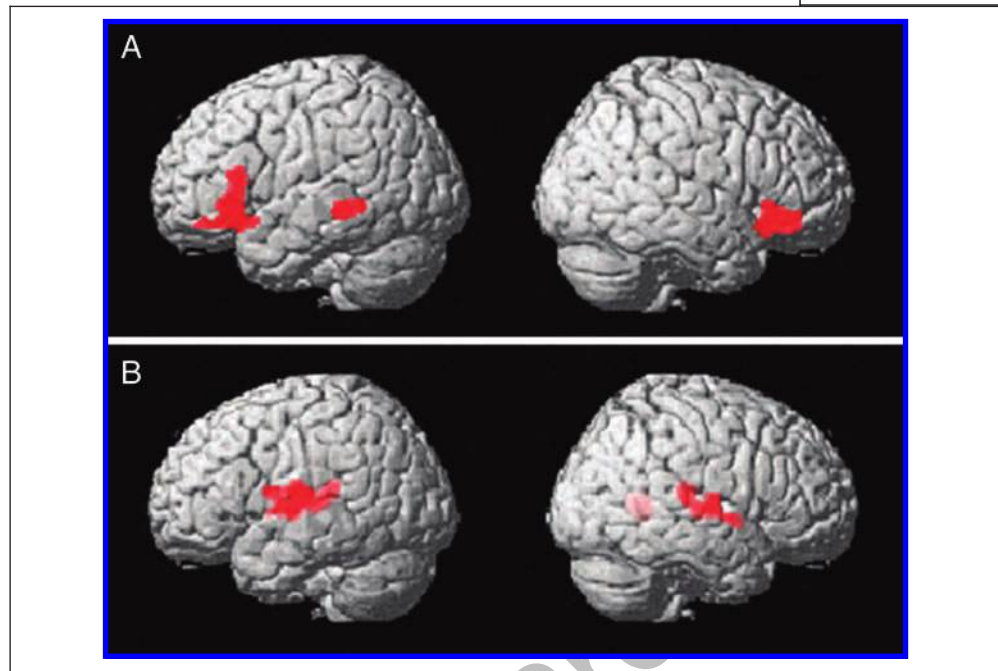


Figure 5A shows the brain regions common to the unification of speaker characteristics, semantic information, and world knowledge.

For the opposite effect, we created a contrast image showing regions involved in correct sentences relative to semantic and world knowledge conditions and then used this image for a small-volume correction on the contrast speaker-congruent > speaker-incongruent sentences. This analysis revealed activation for congruent sentences (i.e., speaker-congruent sentences and sentences without a world knowledge or semantic anomaly) in right PCC [BA 23/31; $t(41) = 4.29, p = .029$]. Figure 5B displays a section with the activation in right PCC.

DISCUSSION

The aim of this fMRI study was to investigate the neural unification of voice-based inferences about speaker characteristics and the lexical content of a spoken sentence. In particular, we wanted to answer the question whether there is an overlap in neural recruitment for unification of core linguistic information and that of extralinguistic, pragmatic information. With respect to these issues, the two main findings of this study are as follows. Firstly, manipulating the congruency of voice-based inferences about the speaker's age, sex, or social background and the semantic content of the spoken sentence showed bilateral involvement of the IFG (BA 45/47) during unification of speaker characteristics and sentence meaning. Secondly, there was an overlap in brain regions involved in the unification of world knowledge, semantic information, and speaker characteristics, thus suggesting a common neural underpinning for the unification of core linguistic and extralinguistic information. Common brain

regions included the IFG bilaterally (BA 45/47) and the posterior part of the left middle temporal gyrus (BA 21).

Inferior Frontal Gyrus and Unification of Speaker Characteristics

As hypothesized, listening to speaker-incongruent sentences increased activation in the IFG (BA 45/47), suggesting that this region is involved in the unification of speaker characteristics and sentence meaning. This result is in line with other findings that suggest a role for inferior frontal cortex in sentence and discourse comprehension (Zempleni et al., 2007; Kuperberg, Lakshmanan, Caplan, & Holcomb, 2006; Rodd et al., 2005; Hagoort et al., 2004; Dapretto & Bookheimer, 1999; Just, Carpenter, Keller, Eddy, & Thulborn, 1996). The involvement of the LIFG in the unification of speaker characteristics inferred from the voice and sentence meaning is consistent with a view of language comprehension in which inferior frontal cortex serves as a core area for unification operations in language (Hagoort, 2005). Owing to the unification contribution of inferior frontal cortex, incoming information is continuously integrated and combined into an unfolding representation of a multiword utterance, such as a sentence. If incoming information is conflicting, as in the case of a mismatch between voice-based speaker inferences and sentence content, the unification load is increased. In the current study, the strongest response to the increased unification load for speaker-incongruent sentences was found in the LIFG. Importantly, as previously reported (Willems et al., 2007; Hagoort et al., 2004) and present in our own data, the observed activation increase in the LIFG for speaker-incongruent

Table 2.

Anatomical Region	BA	Cluster Size	Voxel T Value	MNI Coordinates		
				x	y	z
<i>a. Results from the Whole-brain Analysis for the Contrast World Knowledge Anomaly > Correct Sentences*</i>						
R. Middle/Superior temporal gyrus	21/22	1515	5.92	60	-28	-2
R. STG	38		5.32	56	12	-10
RIFG (pars triangularis)	45		4.79	58	22	4
LIFG (pars triangularis)	45	1185	5.63	-58	20	8
LIFG (pars orbitalis)	47		4.75	-48	34	-4
LIFG (pars triangularis)	45		4.72	-46	30	10
L. Middle frontal gyrus	6	510	5.11	-40	16	42
	6/9		4.92	-42	16	34
	6		4.06	-44	8	40
L. Middle temporal gyrus	21/22	747	4.31	-58	-46	2
	21/22		4.30	-60	-32	0
LIFG	20		3.89	-54	-50	-16
<i>b. Results from the Whole-brain Analysis for the Contrast Correct Sentences > World Knowledge Anomaly*</i>						
R. Cingulate gyrus	24	216	4.58	4	-10	28
	24		3.57	12	-8	40
L. Posterior cingulate gyrus	31/23	358	4.29	10	-32	38
	31		4.12	6	-38	42
	31		4.00	14	-40	36

*Table shows all clusters at a significance level of $p < .05$ corrected at cluster-level (first thresholded at $p < .001$, uncorrected). All local maxima are reported as MNI coordinates. Significant activation peaks > 8 mm apart.

sentences does not reflect a response to a mismatch per se, as the region is also implicated in processing speaker-congruent sentences (see Figure 1). This provides support for the idea that left inferior frontal cortex plays an important role in semantic unification during language comprehension.

Left inferior frontal cortex is a relatively large and anatomically heterogeneous cortical region with numerous

connections to other brain regions. Neuroimaging data suggest that the role of the LIFG extends beyond the language domain and it has been put forward that the main function of the LIFG is “controlled retrieval” or “(semantic) selection” (Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997). Accounts of inferior frontal cortex as playing a key role in the selection of competing semantic

Figure 3. Results of the whole-brain analysis showing clusters activated in response to sentences with world knowledge anomalies relative to correct sentences.

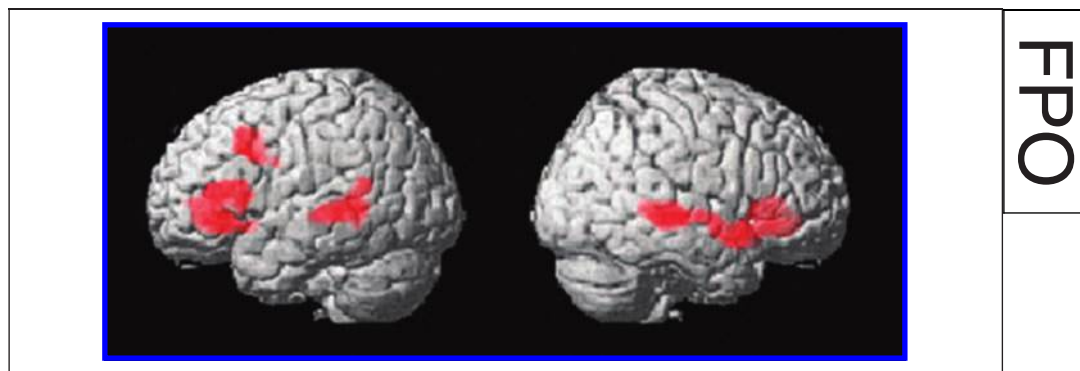


Table 3.

<i>Anatomical Region</i>	<i>BA</i>	<i>Cluster Size</i>	<i>Voxel T Value</i>	<i>MNI Coordinates</i>		
				<i>x</i>	<i>y</i>	<i>z</i>
<i>a. Results from the Whole-brain Analysis for the Contrast Semantic Anomaly > Correct Sentences*</i>						
LIFG (pars triangularis)	45	2187	8.13	-56	20	10
	45		6.98	-52	32	12
	45		6.86	-46	26	14
L. Middle temporal gyrus	21	2030	6.41	-52	-50	4
L. Middle/Superior temporal gyrus	21/22		6.14	-60	-40	6
	21/22		5.71	-58	-30	2
R. Cerebellum		486	5.99	20	-74	-34
			4.00	28	-62	-24
			3.72	42	-68	-30
L. Middle frontal gyrus	9	637	5.71	-50	16	32
L. Frontal operculum	6/9		5.50	-48	8	28
L. Middle frontal gyrus	6		5.18	-38	2	54
R. STG	22	534	4.60	56	-20	-2
	22		4.50	54	-28	-2
	22		4.12	58	-2	-8
<i>b. Results from the Whole-brain Analysis for the Contrast Correct Sentences > Semantic Anomaly*</i>						
R. Inferior rostral gyrus	11	1423	6.75	6	34	-16
Inferior rostral gyrus	10		5.11	0	56	-4
L. Inferior rostral gyrus	12		4.88	-2	22	-20
R. Posterior cingulate gyrus	31	977	6.67	4	-40	40
	31		5.98	10	-46	38
L. Angular gyrus	39/19	445	6.28	-38	-76	36
R. Angular gyrus	39	656	5.93	46	-70	38
	39/19		3.55	34	-78	30

*Table shows all clusters at a significance level of $p < .05$ corrected at cluster-level (first thresholded at $p < .001$, uncorrected). All local maxima are reported as MNI coordinates. Significant activation peaks > 8 mm apart.

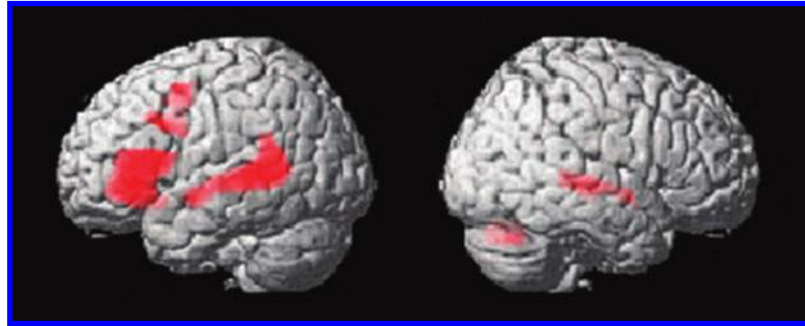
representations are not incompatible with the view of this region as a unification space for language because selection often is an aspect of unification (Vosse & Kempen, 2000). However, it is not clear how the results of this study could be easily explained by selection accounts.

Although speaker-incongruent sentences elicited the strongest BOLD response in the LIFG, they also evoked a significant increase in activation in the homotopic region (BA 47) in the right hemisphere. The observed bilateral activation pattern is compatible with findings from other neuroimaging studies that looked at semantic ambiguity at the sentence level or in a discourse

context (Zempleni et al., 2007; Robertson et al., 2000; St George, Kutas, Martinez, & Sereno, 1999). Recently, it has been suggested that bilateral IFG activation during discourse processing is possibly related to the construction of a situation model (Menenti et al., in press; Ferstl, Rinck, & von Cramon, 2005). A situation model is a mental representation of the situation described by the utterance in connection to preceding or concurrent sources of information (Zwaan & Radvansky, 1998; Van Dijk & Kintsch, 1983).

When encountering information that is implausible or unexpected given the current situation model and general world knowledge, a listener will attempt to revise the

Figure 4. Results of the whole-brain analysis showing clusters activated in response to sentences with semantic anomalies relative to correct sentences.



model by integrating the unexpected information into the ongoing representation of the situation described by the utterance—in-context (Ferstl et al., 2005; Van Dijk & Kintsch, 1983). An fMRI study by Ferstl et al. (2005) suggests that the integration of new or inconsistent information in the situation model involves prefrontal cortex bilaterally (BA 47/11), with slightly more extended activation in the left than in the right IFG (Ferstl et al., 2005). Consequently, the bilateral activation pattern in the IFG observed in our study is consistent with a scenario in which the listener unifies unexpected incoming information and updates the situation model (Nieuwland, Petersson, & Van Berkum, 2007; Ferstl et al., 2005).

The differential contribution of left and right IFG to the unification of incoming information cannot be unraveled by the design of our study. An fMRI study by Menenti et al. (in press) investigated whether manipulating discourse context modulated the unification of

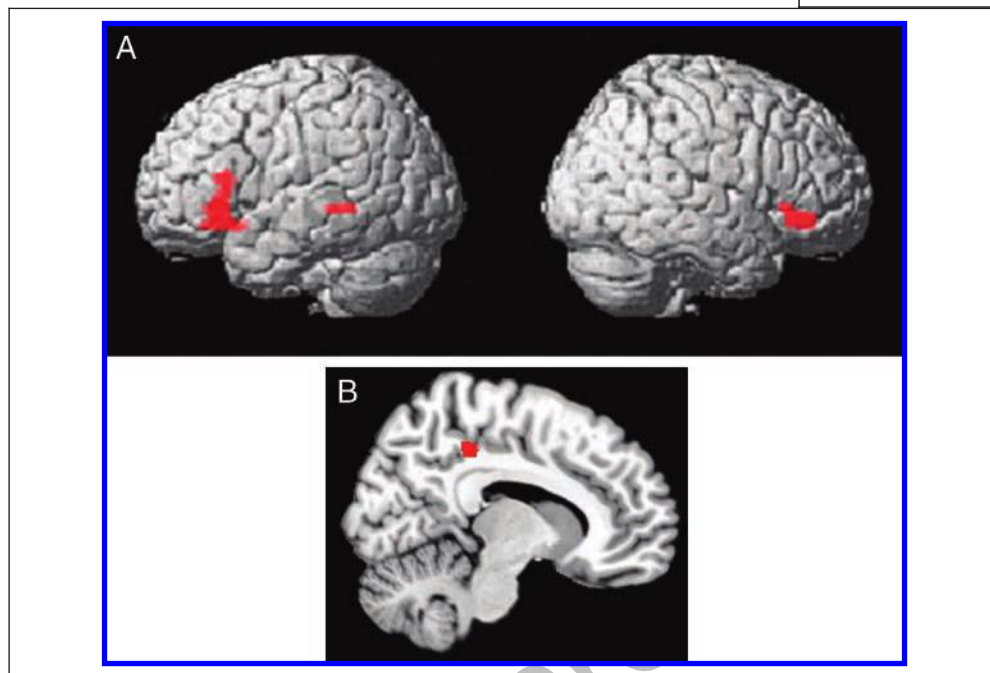
world knowledge. Results showed that the LIFG and RIFG were both recruited in on-line semantic unification of incoming information with previously stored knowledge in long-term memory. Moreover, the LIFG remained sensitive to semantic unification of incoming information with prior world knowledge, even if preceding discourse context overrides this knowledge. In contrast, the RIFG was more sensitive to the local discourse (Menenti et al., in press). These findings suggest a division of labor between the RIFG and the LIFG when it comes to discourse comprehension that might also apply to unification of sentence content with knowledge about the speaker. Several other ideas have been put forward with respect to the precise and possibly different contribution of the two hemispheres in language processing (Mason & Just, 2007; Jung-Beeman, 2005; Faust & Chiarello, 1998), but this issue needs further exploration.

Table 4. Common Regions for Unification of Speaker Characteristics, Semantic Information, and World Knowledge*

<i>Anatomical Region</i>	<i>BA</i>	<i>Cluster Size</i>	<i>Voxel T Value</i>	<i>MNI Coordinates</i>		
				<i>x</i>	<i>y</i>	<i>z</i>
LIFG (pars triangularis)	45	334	4.50	-54	26	14
LIFG (pars triangularis/pars orbitalis)	45/47		4.22	-48	26	-2
LIFG (pars triangularis)	45		4.07	-50	22	8
LIFG (pars orbitalis)	47		4.04	-48	24	-6
	47		3.81	-46	26	-12
	47		3.76	-46	34	-12
L. Temporal pole	38		3.69	-52	16	-12
LIFG (pars orbitalis)	47		3.51	-38	22	-12
RIFG (pars orbitalis)	47	104	4.48	50	34	-12
	47		4.40	50	40	-12
	47		4.30	54	28	-6
L. Middle temporal gyrus	21	74	4.18	-58	-42	-4
	21		3.95	-62	-36	-4

*Table shows all clusters at a significance level of $p < .05$ corrected at cluster-level (first thresholded at $p < .001$, uncorrected). All local maxima are reported as MNI coordinates. Significant activation peaks > 4 mm apart.

Figure 5. (A) Common regions of activation for unification of linguistic (world knowledge and semantic knowledge) and extralinguistic information (speaker characteristics). (B) Common regions of activation for processing congruent sentences (i.e., speaker-congruent sentences and sentences without a semantic or world knowledge anomaly).



Common Neural Correlates for Unification of Speaker Characteristics, Semantic Information, and World Knowledge

Next to examining the neural correlates of unification of speaker characteristics and sentence meaning, we wanted to identify brain regions that are involved in unifying both core linguistic information and extralinguistic information. Common neural correlates for unification of world knowledge, semantic information, and speaker characteristics were the IFG bilaterally (BA 45/47) and the posterior part of the left middle temporal gyrus (BA 21). The observed overlap in cortical regions points to similarities in neural recruitment for the unification of linguistic and extralinguistic information. Further evidence for the resemblance in unification comes from the finding by Van Berkum et al. (2008) that semantic and world knowledge anomalies elicit an ERP effect, the so-called N400 effect, with the same temporal and spatial distribution as speaker-incongruent sentences. The importance of the IFG for unification has been discussed above. The left middle temporal gyrus plays a key role in the storage and retrieval of semantic information (Hickok & Poeppel, 2007; Hagoort, 2005; Indefrey & Cutler, 2005). Functional neuroimaging studies using semantically ambiguous sentences have shown that, in particular, posterior middle temporal regions are important for lexical-semantic processing (Zemleni et al., 2007; Rodd et al., 2005).

According to our view, the left posterior middle temporal gyrus (LpMTG) can be considered as a component of the unification network that also involves the LIFG and the RIFG. Within the unification network, there possibly is a dynamic interplay between inferior frontal

regions and the LpMTG that explains the observed activation pattern in our study (see also Snijders et al., 2006). Speculatively, this interaction between frontal and posterior temporal cortex would serve to maintain the retrieved semantic information on-line so that unification can take place. Here, it is important to note that the anomalous sentences in our experiment did not contain a violation in a strict sense. As a consequence, unification is not precluded, but the processing of anomalous sentences is associated with an increased unification load that needs prolonged activation of semantic information. This is, for example, needed for our speaker-incongruent sentences, where it is harder to associate the critical word with the voice of the speaker and thereby the unification load is increased. For unification to be achieved, it is thus required that the semantic information retrieved for the critical word remains activated. It is known that frontal cortex exerts top-down control over more posterior regions, such as the LpMTG, where (semantic) representations are stored (Curtis & D'Esposito, 2003; Miller & Cohen, 2001). These feedback signals from inferior frontal cortex influence which information is maintained by posterior areas and might, in our case, make sure that relevant semantic representations remain active so that unification can take place.

Support for such a left frontal-temporal interplay comes from a study by Kuperberg et al. (2003), in which pragmatically anomalous (comparable to the world knowledge anomalies in our study), morphosyntactically anomalous, and correct sentences were presented. fMRI results showed that the same regions within a left temporal-frontal network were modulated to different degrees by both pragmatically and morphosyntactically anomalous sentences. Combining these fMRI results with reaction

time (RT) data from plausibility ratings of the sentences revealed that the pattern of response within the left temporal–frontal network across the three sentence types mirrored the pattern of RTs, with most activity and the longest RTs in association with the pragmatically anomalous sentences, and least activity and the shortest RTs in association with the morphosyntactically anomalous sentences. Increased neural activity in the left temporal–frontal network, together with longer RTs for the pragmatically anomalous sentences, was interpreted as reflecting increased and more prolonged efforts to search and retrieve semantic knowledge about the likelihood of events occurring in the real world (see also Kuperberg, Sitnikova, & Lakshmanan, 2008; Kuperberg et al., 2003). It was suggested that in the case of morphosyntactically anomalous sentences, RTs and neural activity were reduced because plausibility decisions about these sentences can be made on the basis of a finite set of syntactic rules.

We also examined the reversed contrast to determine whether there was an overlap in brain regions recruited for congruent sentences (i.e., speaker-congruent sentences and sentences without a semantic or world knowledge anomaly) relative to speaker-incongruent sentences and sentences with a world knowledge or semantic anomaly. This revealed a significant cluster of activation in right PCC (BA 23/31). A meta-analysis by Ferstl, Neumann, Bogler, and von Cramon (2008) has shown that the left hemisphere counterpart of this region is activated for comprehending coherent language compared with incoherent language, suggesting that this region is important for coherence building. Stronger activity in right PCC observed for congruent sentences compared to sentences with an anomaly might be related to a role for this region in processing coherent language. More in general, the observed activation for congruent sentences might be related to PCC being part of the default mode network (Buckner, Andrews-Hanna, & Schacter, 2008; Raichle et al., 2001). Activity in default mode brain regions (among which PCC) is attenuated as a function of task difficulty (Greicius & Menon, 2004; Gusnard, Raichle, & Raichle, 2001; Raichle et al., 2001). In our study, congruent sentences are less attention demanding and engaging than sentences with a speaker, semantic, or world knowledge anomaly, and will therefore cause less suppression of activity in default mode regions (i.e., in PCC; see also Wilson, Molnar-Szakacs, & Iacoboni, 2008). Thus, it is plausible that stronger activation in right PCC for congruent sentences reflects less disrupted default activity for these sentences than for sentences with an anomaly.

Brain Regions Involved in Processing Speaker-congruent Sentences

The contrast speaker-congruent versus speaker-incongruent sentences showed stronger activation for congruent sentences bilaterally in Heschl's gyrus and the STG, with

greater activation on the left extending to the planum temporale. In addition, there was increased activation in the right lingual gyrus that extended into PCC. The bilateral activation in superior temporal cortex (BA 22/41/42) is consistent with findings from studies on speech perception and auditory sentence comprehension (Constable et al., 2004; Hickok & Poeppel, 2000) that show stronger involvement of this region during processing of semantically meaningful (coherent) and intelligible speech (Humphries, Binder, Medler, & Liebenthal, 2006; Davis & Johnsrude, 2003). The fit between voice-based inferences and sentence content is what makes speaker-congruent sentences easier to process than speaker-incongruent sentences. Given the selective sensitivity of superior temporal regions to coherence in linguistic information and their special role in voice processing (Belin et al., 2000), the activation in these regions for speaker-congruent sentences is possibly due to successful support of voice-based inferences in line with the whole sentence meaning. Importantly, the observed superior temporal cortex activation seems specific for the congruence between voice and message and is unlikely to be due to coherence in general. This follows from the fact that this region does not show up as significantly activated in the abovementioned contrast of correct sentences versus sentences with a speaker, semantic, or world knowledge anomaly.

Conclusion

In conclusion, the present study not only replicates earlier findings on the integrative role of left inferior frontal cortex in language comprehension but it also extends the existing knowledge about the nature of the information that is unified. From fMRI studies on unification during visual and auditory sentence comprehension, we know that, within the language domain, the involvement of left inferior frontal cortex in unification processes is independent of input modality (i.e., it operates during reading as well as during understanding speech; Willems et al., 2007; Hagoort et al., 2004). Furthermore, findings from an fMRI study by Willems et al. (2007) on unification of cospeech gestures have shown that unification space in the LIFG is not domain specific: It integrates semantic information coming from the speech domain as well as from the action domain (i.e., as extracted from gestures). In our study, the information that needed to be unified had yet a different nature. Both sources of information (sentence meaning and speaker characteristics) came from the same modality because they were both extracted from the speech signal. However, they differed in another dimension. Although sentence meaning per se is semantic in nature, voice-based inferences about characteristics of the speaker can be regarded as more pragmatic and also social in nature. Thus, our findings suggest that the role of the LIFG is not exclusively limited to the unification of language

information, but it also plays a significant role in the on-line unification of extralinguistic information, including social information concerning speaker characteristics carried by the speech signal. In short, the data suggest that the LIFG unifies multiple sources of information during language comprehension, linguistic as well as extralinguistic. Finally, we identified an overlap in brain regions involved in unifying speaker characteristics, semantic information, and world knowledge. This further confirms that unification processes for core linguistic and extralinguistic information have shared underlying neural correlates and that during language comprehension not only information from a broad range of cognitive domains is incorporated, but information from the social domain is also taken into account.

Reprint requests should be sent to Cathelijne M. J. Y. Tesink or Peter Hagoort, F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands, or via e-mail: c.tesink@fcdonders.ru.nl; p.hagoort@fcdonders.ru.nl.

REFERENCES

- Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., & Zilles, K. (2000). Brodmann's areas 17 and 18 brought into stereotaxic space—Where and how variable? *Neuroimage*, *11*, 66–84.
- Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron*, *47*, 907–918.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135.
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, *14*, 2105–2109.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649–657.
- Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., et al. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *Neuroimage*, *22*, 11–21.
- Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*, 415–423.
- Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, *24*, 427–432.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*, 3423–3431.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335.
- Faust, M., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. *Neuropsychologia*, *36*, 827–835.
- Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex?—An fMRI study. *Neuroimage*, *23*, 840–848.
- Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, *29*, 581–593.
- Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *17*, 724–739.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.
- Greicius, M. D., & Menon, V. (2004). Default-mode activity during a passive sensory task: Uncoupled from deactivation but impacting activation. *Journal of Cognitive Neuroscience*, *16*, 1484–1492.
- Gusnard, D. A., Raichle, M. E., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*, 685–694.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*, 416–423.
- Hagoort, P., Baggio, G., & Willems, R. M. (in press). Semantic unification. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences*. Cambridge, MA: MIT Press.
- Hagoort, P., & Brown, C. M. (1994). Brain responses to lexical-ambiguity resolution and parsing. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 45–80). Hillsdale, NJ: Erlbaum.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131–138.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, *18*, 665–679.
- Indefrey, P., & Cutler, A. (2005). Prelexical and lexical processing in listening. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 759–774). Cambridge, MA: MIT Press.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, *1146*, 2–22.
- Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *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.

- 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. *Journal of Cognitive Neuroscience*, *15*, 272–293.
- Kuperberg, G. R., Lakshmanan, B. M., Caplan, D. N., & Holcomb, P. J. (2006). Making sense of discourse: An fMRI study of causal inferencing across sentences. *Neuroimage*, *33*, 343–361.
- Kuperberg, G. R., Sitnikova, T., & Lakshmanan, B. M. (2008). Neuroanatomical distinctions within the semantic system during sentence comprehension: Evidence from functional magnetic resonance imaging. *Neuroimage*, *40*, 367–388.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131.
- Mai, J. K., Assheuer, J., & Paxinos, G. (2004). *Atlas of the human brain*. London: Elsevier.
- Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, *1146*, 115–127.
- Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (in press). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience*.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience*, *12*, 120–133.
- Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *Neuroimage*, *37*, 993–1004.
- Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, *28*, 383–407.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 676–682.
- Robertson, D. A., Gernsbacher, M. A., Guidotti, S. J., Robertson, R. R., Irwin, W., Mock, B. J., et al. (2000). Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychological Science*, *11*, 255–260.
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*, 1261–1269.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2006). *From words to sentences: Retrieval and unification processes in the brain*. Poster presented at the 12th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Nijmegen, The Netherlands.
- St George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain*, *122*, 1317–1325.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, U.S.A.*, *94*, 14792–14797.
- van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*, 962–974.
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, *17*, 48–55.
- von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*, 948–955.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.
- Willems, R. M., Ozyurek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*, 2322–2333.
- Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2008). Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cerebral Cortex*, *18*, 230–242.
- 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.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.