



Differential contributions of left-hemispheric language regions to basic semantic composition

Astrid Graessner^{1,2} · Emiliano Zaccarella² · Gesa Hartwigsen^{1,2}

Received: 5 August 2020 / Accepted: 16 December 2020 / Published online: 30 January 2021
© The Author(s) 2021

Abstract

Semantic composition, the ability to combine single words to form complex meanings, is a core feature of human language. Despite growing interest in the basis of semantic composition, the neural correlates and the interaction of regions within this network remain a matter of debate. We designed a well-controlled two-word fMRI paradigm in which phrases only differed along the semantic dimension while keeping syntactic information alike. Healthy participants listened to meaningful (“fresh apple”), anomalous (“awake apple”) and pseudoword phrases (“awake gufel”) while performing an implicit and an explicit semantic task. We identified neural signatures for distinct processes during basic semantic composition. When lexical information is kept constant across conditions and the evaluation of phrasal plausibility is examined (meaningful vs. anomalous phrases), a small set of mostly left-hemispheric semantic regions, including the anterior part of the left angular gyrus, is found active. Conversely, when the load of lexical information—independently of phrasal plausibility—is varied (meaningful or anomalous vs. pseudoword phrases), conceptual combination involves a wide-spread left-hemispheric network comprising executive semantic control regions and general conceptual representation regions. Within this network, the functional coupling between the left anterior inferior frontal gyrus, the bilateral pre-supplementary motor area and the posterior angular gyrus specifically increases for meaningful phrases relative to pseudoword phrases. Stronger effects in the explicit task further suggest task-dependent neural recruitment. Overall, we provide a separation between distinct nodes of the semantic network, whose functional contributions depend on the type of compositional process under analysis.

Keywords Meaning composition · Angular gyrus · Conceptual combination · fMRI · Functional connectivity

Introduction

One of the core features of human language is the ability to combine single words into complex phrases. Semantic knowledge helps us to make sense of words and semantic composition processes drive the way we combine individual meanings into more composite ones. Current neuroanatomical models of semantic processing highlight a widely

distributed fronto-temporo-parietal network in the left hemisphere (Binder et al. 2009). Focusing on semantic composition, a number of studies have identified several brain regions in the left hemisphere showing higher activation for sentences than word lists, including the angular gyrus (AG) in the inferior parietal lobe, posterior middle temporal gyrus (pMTG), anterior temporal lobe (ATL) and the anterior inferior frontal gyrus (aIFG) (Brennan and Pykkänen, 2012; Humphries et al. 2006; Lau et al. 2008; Matchin et al. 2017; Pallier et al. 2011; Vandenberghe et al. 2002; Vigneau et al. 2006). However, due to the complexity of sentential manipulations, it can be difficult to differentiate processes of semantic composition from those including other linguistic operations such as syntax and other cognitive domains such as working memory, attention and cognitive control (Badre 2008; Makuuchi and Friederici 2013). In recent years, neuroscientific researchers have become increasingly more interested in simpler paradigms using two- or three-word

Emiliano Zaccarella and Gesa Hartwigsen shared last authorship.

✉ Astrid Graessner
graessner@cbs.mpg.de

¹ Lise-Meitner Research Group Cognition and Plasticity, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

² Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

phrases to tackle semantic composition in more controlled linguistic constructions (Pylkkänen 2019).

One of the proposed key regions for basic semantic composition is the AG. Several studies report recruitment of the left (and sometimes right) AG for two-word phrases relative to single words (Bemis and Pylkkänen 2013b), meaningful as compared to meaningless adjective-noun phrases (Graves et al. 2010; Molinaro et al. 2015; Price et al. 2015) and for tracking thematic relations between words (Boylan et al. 2015, 2017; Lewis et al. 2018). Note, however, that in an alternative account, the AG is interpreted to belong to the default mode network (DMN) and its contribution to semantic tasks may reflect less deactivation for easier task conditions (Humphreys et al. 2015; Lambon Ralph et al. 2016).

Another region that has consistently been implicated as a key semantic composition region is the ATL. Within the “hub-and-spokes” model, the ATL is considered to bind information from different modalities (the “spokes”), serving as a transmodal “hub” (see Lambon Ralph et al. 2016). This view is supported by the finding that patients with semantic dementia who show bilateral atrophy of the ATL are impaired in semantic processing across all input modalities and types of concepts (Mummery et al. 2000). Regarding basic semantic composition, most evidence for the ATL as the key region of conceptual combination is derived from Magnetoencephalography (MEG) studies. Increased activity in the bilateral ATL has been shown for two-word phrases as compared to single words in both the visual and auditory modality (Bemis and Pylkkänen 2013b) and in different languages, including English, Spanish, Arabic, and American Sign Language (Bemis and Pylkkänen 2011; Blanco-Elorrieta et al. 2018; Molinaro et al. 2015; Westerlund et al. 2015).

Functional MRI evidence for the ATL as a key region for basic semantic composition is rather scarce, which might be explained by several methodological issues. First, the ATL suffers from signal loss in fMRI due to its location near the sinuses. Second, the use of low-level baselines which still engage semantic processing (e.g., internal speech) might have led to a lower likelihood of finding ATL activation because semantic activation in this region was removed during the subtraction analyses (Visser et al. 2010).

Finally, a region that has received less attention in the context of semantic composition but has rather been referred to as an executive semantic control region is the left aIFG (BA45/47). Increased aIFG activation has been observed in previous studies for sentences compared to word lists (Matchin et al. 2019a, b; Pallier et al. 2011), ambiguous relative to unambiguous sentences (Rodd et al. 2005; Vitello et al. 2014) and two-word phrases compared to single words (Schell et al. 2017). The latter authors emphasize that during basic semantic composition, BA45 increases its activation level with the amount of words that can be integrated into

context. Thus, two-word phrases should always elicit higher activity in left aIFG than single words.

Despite considerable effort to characterize the neural correlates for basic semantic composition, several questions remain open. First, although there is consensus that left AG, ATL and aIFG play important roles in semantic composition, most fMRI studies have focused on the contribution of single regions instead of investigating functional interactions at a larger network level. Consequently, it remains unclear how these regions influence each other during semantic composition. At the single-word level, Hartwigsen et al. (2015) have shown that left AG and aIFG were able to compensate for a focal disruption of the respective other region induced by transcranial magnetic stimulation (TMS). In that study, semantic decisions were only impaired after both regions had been perturbed, showing that the interplay of these regions is causally relevant for semantic decisions. However, it is not clear whether this interaction is restricted to the single-word level or whether it is also involved in compositional processing.

Furthermore, it remains unclear whether the previously reported recruitment of brain areas during semantic composition is task-dependent. While some studies explicitly asked subjects to compose the meaning of the stimuli (Bemis and Pylkkänen 2011; Graves et al. 2010; Price et al. 2015; Schell et al. 2017), others intentionally did not (Graves et al. 2010; Matchin, Brodbeck et al. 2019a, b; Matchin et al. 2017; Matchin et al. 2019a, b; Molinaro et al. 2015). Few studies have directly compared different tasks while keeping the stimulus material similar and different results between studies might thus reflect differences in task processing and demand.

Finally, we note that the few existing two-word studies have used different baseline conditions. While some studies compared the processing of two-word phrases to single words (e.g., Bemis and Pylkkänen 2011, 2013a; Schell et al. 2017), others looked at more meaningful versus less meaningful two-word combinations (Graves et al. 2010; Molinaro et al. 2015; Price et al. 2015). It is conceivable that the different semantic regions discussed above fulfill distinct tasks in combining the meaning of two separate concepts into a whole.

To address this issue, it is thus advantageous to investigate the potential different processes in one single experimental setting. In the current experiment, we created a paradigm consisting of three different two-word phrases: meaningful phrases (“fresh apple”), anomalous phrases (“awake apple”), and pseudoword phrases containing pseudonouns (“awake gufel”; see “Methods” section for details).

This design allowed us to first measure two *specific processes* directly tackling semantic plausibility: (1) *meaningful composition* (meaningful > anomalous) and (2) *anomalous composition* (anomalous > meaningful),

where the two conditions only differ with regard to plausibility. Second, given that the syntactic information was kept constant across conditions, while the load of lexical information was manipulated, we could separate (3) *meaningful phrasal composition* (meaningful > pseudoword phrase) from (4) *anomalous phrasal composition* (anomalous > pseudoword phrase). Additionally, the overlap of the latter two contrasts would show regions that are involved in (5) *general phrasal composition*, independent of the plausibility of the resulting phrase. Furthermore, to explore whether the activation of semantic core regions is task-dependent, we included both an implicit and an explicit semantic task with respect to phrasal plausibility. This allowed us to distinguish task-specific and automatic processes during semantic composition. If semantic composition occurs implicitly and independent of task demands, we should see similar results for the explicit and implicit tasks.

Finally, we conducted psychophysiological interaction analyses to assess task-related changes in functional connectivity within the semantic composition network. This network perspective is receiving increasing interest in the study of neurocognitive processes as it enables to better characterize how the brain is organized along large scale networks (Bassett and Sporns 2017; Hartwigsen 2018). In summary, we aimed to provide a comprehensive characterization of the network for basic semantic composition and explore potential task dependencies in the activation patterns.

Based on the above-cited studies, we expected to find strong involvement of the left AG, ATL and aIFG for the contrast of meaningful > pseudoword phrases, reflecting increased load of lexical information during the formation of complex meaning. Within this network, the AG should be more specifically recruited during the processing of meaningful phrases compared to anomalous phrases (*meaningful composition*), as an effect of semantic plausibility. Conversely, the aIFG should be maximally recruited during the processing of anomalous phrases compared to meaningful phrases (*anomalous composition*), as a function of higher semantic control. By administering both an implicit and an explicit task, we aimed to identify regions that are activated in a task-dependent manner. We expected more inferior frontal involvement for the explicit task, while we hypothesized regions that are associated with automatic semantic processing (e.g. left ATL, AG) to be activated also in the implicit task. A potential overlap in activation for the two tasks was expected to reflect task-independent semantic composition processes. Regarding functional connectivity, we hypothesized that an interaction between left AG and aIFG, as previously observed during single word processing, would drive the comprehension of two-word phrases. However, we were also interested in the connectivity of other semantic network

regions and therefore conducted several exploratory analyses with seed regions from the univariate GLM results.

Methods

Participants

Thirty-seven right-handed German-speaking subjects participated in the two sessions of the study. They had normal hearing, corrected to normal vision and no history of neurological disorders or contraindication to MR-scanning. Four participants had to be removed from the analyses due to low task accuracy (see Behavioral Analysis). The final group of participants that entered the analyses consisted of 33 participants (16 females, mean age 26 years, $SD = 3.6$ years).

All participants gave their written informed consent and were reimbursed with 10€/hour. The study protocol conformed to the principles of the Declaration of Helsinki and was approved by the local ethics committee at the University of Leipzig.

Experimental paradigm

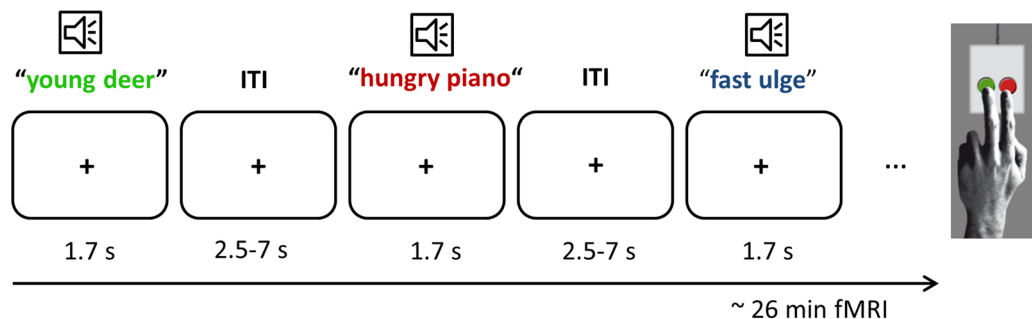
All participants completed two event-related fMRI sessions separated by at least one week with two different tasks on the same set of stimuli. Auditory stimuli were presented using the software package Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA) via MR-compatible in-ear headphones (MR-Confon, Magdeburg, Germany). Volume was adjusted to an optimal individual hearing level. Stimuli consisted of spoken word pairs that were either meaningful (“fresh apple”), anomalous (“awake apple”) or had the noun replaced by a pseudoword (“awake gufel”) (see Fig. 1a).

In the first session, participants performed an *implicit* task with respect to the meaning of the phrases. They were instructed to indicate whether both words had the *same* lexical status (i.e., both real words or both pseudowords) or whether they had a *different* lexical status (i.e., one real word and one pseudoword). Based on the vast literature on priming (cf. Lau et al. 2008), we expected our altered lexical decision task to capture automatic semantic processes. Please note that this implicit task was designed to match task demands to the explicit task and thus requires a higher amount of cognitive effort than classical implicit tasks (e.g. lexical decision). We nonetheless refer to it as implicit as it stays implicit regarding the compositional meaning of the phrase. We added filler trials with two pseudowords and a pseudo-adjective paired with a real noun, to prevent participants from focusing on the second word only. In addition, single word trials served as control stimuli to balance the positive and negative responses. Participants were trained that single words required a “*different*” response, as there

a Experimental Paradigm

Condition	Example	Task
Meaningful	<i>frischer Apfel</i> (<i>fresh apple</i>)	Implicit: Comparison of lexical status of the words Explicit: Meaningfulness judgement of the phrase
Anomalous	<i>wacher Apfel</i> (<i>awake apple</i>)	
Pseudoword	<i>frischer Gufel</i> (<i>fresh gufel</i>)	

b Example of three trials



c Contrasts

Meaningful composition

fresh apple > awake apple
(Meaningful > Anomalous)

Meaningful phrasal composition

fresh apple > fresh gufel
(Meaningful > Pseudoword)

Anomalous composition

awake apple > fresh apple
(Anomalous > Meaningful)

Anomalous phrasal composition

awake apple > fresh gufel
(Anomalous > Pseudoword)

Fig. 1 Experimental Design. **a** Experimental conditions and task descriptions used in the study. **b** Example of three trials. The inter-trial-interval (ITI) was jittered from 2500 ms up to 7000 ms with a mean duration of 4000 ms. **c** Contrasts of interest with involved compositional processes

was no other word to compare it to. We note that this might not have been the intuitive response and made the task more difficult, however, we did not include the single word trials in any direct comparisons, as they served only as control trials.

In the second session, participants performed an *explicit* meaningfulness judgement task by indicating whether the phrase they heard was meaningful or not. To keep the amount of positive and negative responses equal, similarly to the implicit task, we additionally added single (meaningful) nouns as fillers which were not analyzed. Again, participants were trained to respond with “*meaningful*” to the single word condition. Before each session, participants completed a practice session outside of the scanner with a separate stimulus set. Subjects gave their response via button press of the left index or middle finger. Response button assignment was counterbalanced across participants. Both

sessions consisted of 8 blocks with all conditions appearing 7 times in each block and pseudo-randomized with respect to order across participants. Blocks were separated by 20-s rest periods. Note that the implicit task was always performed in the first session, to keep the processing implicit and not biased by a previous task. The overall length of the two sessions was 31 min for the implicit task and 26 min for the explicit task (more filler trials in the implicit task, see *Stimulus* section).

Stimuli

With our design, we aimed to isolate neural responses during the process of semantic composition, removing any effect of syntactic processing. To this end, we created a paradigm using one meaningful composition condition (*meaningful*, e.g. “fresh apple”) and two conditions where the composition

is difficult or even impossible: In the *anomalous* condition, we created real word phrases that could not be combined using world knowledge (e.g. “awake apple”). In these stimuli, the adjective violated the selectional restriction criteria of the noun. The adjective “awake” typically describes living things but cannot be mapped onto non-living entities such as apples. For more types of violations, a list of all stimuli can be found in SI Table 2. In the *pseudoword* condition, we replaced the noun with a pseudonoun but kept the adjective the same. This way, the syntactic information is comparable for the pseudoword condition, but as it only contains one real word, a combination of concepts is not possible. This provided us with the advantage of avoiding confounds in the number of words presented and instead reduced the amount of conceptual information specifically.

To create real word stimulus pairs (*meaningful* and *anomalous*), we selected 400 nouns from the SUBTLEX-DE database (Brysbaert et al. 2011), constraining our search to the following criteria: mono- or disyllabic, masculine or neuter, monomorphemic, concrete, mean log frequency of 2.56. Concreteness was determined using an English corpus with ratings for 40,000 words (Brysbaert et al. 2014). As there is no existing large database for German, we translated the German corpus LANG (Kanske and Kotz 2010) consisting of 1000 nouns into English and correlated the concreteness measures with each other. The correlation was very high between the two databases ($r=0.87$, $p<0.001$) and we thus used the English norms for our German words.

In the next step, we excluded all nouns that were ambiguous in their meaning. Orthographic neighborhood was controlled by calculating the Orthographic Levenshtein Distance 20 (OLD20, R package ‘vwr’), excluding all words deviating more than 1.5 times the interquartile range (IQR) from the mean OLD20 value across all items. Using the same parameters (frequency, number of syllables, OLD 20), except for the concreteness value specific to nouns, we selected adjectives that could modify concrete nouns. Even though we note that using adjectives which normally modify abstract nouns could have easily resulted in anomalous phrases, we decided against using those, to avoid having a confound of activation differences coming from the concreteness value.

To create meaningful word pairs, we generated all possible combinations of adjectives and nouns and assessed each pair’s frequency in the google web1t database (Linguistic Data Consortium, University of Pennsylvania). This database consists of n-gram counts of approximately 100 billion word tokens and searches for the frequency of each word pair. We excluded metaphoric pairs, alliterations and those that deviated more than 1.5*IQR from the mean pair frequency. We then created anomalous pairs by combining adjectives with nouns that did not occur in the google web1t output, taking care that each adjective combined with at least one noun meaningfully and anomalously and that the same

held true for the nouns. In a pilot study, 20 participants who did not take part in the main experiment rated the plausibility of the remaining word pairs on a Likert scale from 1–6. We additionally asked for any potential associations to filter out items that could be understood metaphorically. From these ratings, we selected the highest rated pairs for the meaningful condition and the lowest rated pairs for the anomalous condition, excluding pairs that deviated more than 1.5*IQR from the mean of each condition. The final set of stimuli consisted of 56 phrases per condition. A list of psycholinguistic variables for both conditions can be found in Supplementary Table 1.

To create pseudowords that were comparable to real words, we used the pseudoword generator software Wuggy (Keuleers and Brysbaert 2010). Here, we included all final real word stimuli and matched the pseudowords for the length of subsyllabic segments, letter length, transition frequency and selected those items that deviated the least from the original words in their OLD20 value. Pseudo-homophones (i.e., stimuli that were pronounced as real words) were excluded.

Stimuli were recorded by a professional male speaker in a sound-attenuating chamber with a resolution of 16 bits and a sampling rate of 44.1 kHz. All words were spoken individually in the form of a statement. Thereafter, we cut all words into single files and normalized them to root-mean-square amplitude using the Praat software (version 6.0.04).

To keep the length of stimuli comparable between two-word and single-word phrases, we concatenated word-pairs with a constant noun onset at 1.1 s and embedded all stimuli in speech-shaped noise with a signal to noise ratio of 30. The pause between the words was always 40 ms, ensuring a natural-sounding phrase.

We furthermore included pseudoword-pseudoword and pseudoword-real-word pairs in the implicit task to ensure that participants could not make their judgment based on the second word only. Single nouns additionally served as fillers to keep the amount of positive and negative responses equal in both tasks. Fillers were not analyzed in the subsequent phases. The final set of stimuli had a mean length of 1703 ms (SD = 92 ms).

fMRI acquisition

Functional images were acquired with a 3 T Siemens Magnetom Scanner (Siemens, Erlangen) using a 32-channel head coil. To guarantee optimal signal of the ATL regions (Halai et al. 2014), we adopted a multiband dual gradient-echo echo-planar imaging (EPI) sequence (60 slices in the axial direction and interleaved order, TR = 2 s, short TE = 12 ms, long TE = 33 ms, flip angle of 90°, FOV = 204, slice thickness = 2.5 mm, interslice gap = 0.25 mm, multiband acceleration factor = 2) (Feinberg et al. 2010; Moeller et al. 2010).

To further decrease artifacts in the ATL, slices were tilted by 10° off the AC-PC line. For offline distortion correction, field maps were acquired using a gradient dual-echo sequence (TR = 620 ms, TE1 = 4 ms and TE2 = 6.46 ms). Structural T1-weighted images were previously acquired and retrieved from the institute brain database for all participants. Images were acquired using an MPRAGE sequence (176 slices in sagittal orientation; TR: 2.3 s; TE: 2.98 ms; FoV: 256 mm; voxel size: $1 \times 1 \times 1$ mm; no slice gap; flip angle: 9° ; phase encoding direction: A/P).

Behavioral analysis

Analysis of the behavioral data was performed using the software R (Version 3.2.3). We calculated the mean percentage of correctly answered trials per participant and excluded any participant who performed with less than 75% across all main conditions in any of the sessions (2 participants per session). For the analysis of reaction times, we only considered correctly answered trials within a response-time cutoff range of 2500 ms. All reaction times that deviated more than 3 SD from the mean per participant and condition were excluded (1.9% in the explicit task, 1.8% in the implicit task).

Statistical analyses were performed with the generalized linear mixed-effects model (GLMEM) using the lme4 package in R (Bates et al. 2014), assuming a Gamma distribution of our reaction time data. For the analysis of accuracy, we computed a mixed logit regression. We included by-participant intercepts to account for overall inter-individual differences and by-item intercepts and calculated two models with the respective reference levels *meaningful* and *anomalous*.

fMRI analysis

fMRI analyses were performed using SPM12 (Wellcome Trust Centre for Neuroimaging, <http://www.fil.ion.ucl.ac.uk/spm/>). The functional images from the two echoes were combined using a custom Matlab script that combined the images using a mean weighting by the temporal signal-to-noise ratio (tSNR) at each voxel. The combined functional images were then realigned to the first image, distortion corrected (using the field maps), co-registered to their corresponding structural image, normalized to MNI space (using a unified segmentation with a resampling size of 2.5 mm isotropic voxels) and smoothed with a 5 mm^3 FWHM Gaussian kernel.

For statistical analyses, we estimated a general linear model (GLM) for each participant as implemented in SPM12, including one regressor for each condition and convolving the onset and duration of stimulus presentation with a canonical hemodynamic response function (HRF). Only correctly answered trials were analyzed and we added error

trials as a regressor-of-no-interest. The 6 motion parameters were treated as nuisance regressors. A high-pass filter with 128 s cutoff was applied. For each subject we estimated the contrast for each condition against rest as well as direct contrasts between conditions (meaningful > anomalous, meaningful > pseudowords, anomalous > meaningful, anomalous > pseudowords). At the group level, we conducted one-sample t-tests within each task (implicit and explicit) for each direct contrast. To identify brain regions that are involved in the *specific* processes for basic composition, we contrasted *meaningful* versus *anomalous* phrases (*meaningful composition*) and *anomalous* versus *meaningful* phrases (*anomalous composition*; see Fig. 1c).

We further contrasted *meaningful* versus *pseudoword* phrases and *anomalous* against *pseudoword* phrases to identify brain regions that are engaged for more complex meanings compared to simpler ones.

To explore regions that were activated independently of the final meaningfulness in the explicit task, we performed a conjunction analysis of *meaningful* > *pseudoword* and *anomalous* > *pseudoword* phrases based on the minimum statistic (Nichols et al. 2005), resulting in *general phrasal composition*. Furthermore, to detect regions that were activated independently of the task, we performed a conjunction analysis for the contrasts in both tasks.

Finally, to localize brain regions which responded significantly more to the explicit than the implicit task, we conducted paired t-tests. We, therefore, subtracted the contrast resulting from the implicit task from the one in the explicit, e.g., ($\text{meaningful}_{\text{explicit}} > \text{anomalous}_{\text{explicit}}$) – ($\text{meaningful}_{\text{implicit}} > \text{anomalous}_{\text{implicit}}$). These interactions were inclusively masked by the significant voxels of the minimum to restrict them to those voxels that were also activated in the task (cf. Hardwick et al. 2018).

All contrasts were thresholded using a voxel-wise false discovery rate (FDR) correction of $q < 0.05$ with a cluster-extent threshold of 20 voxels to avoid reporting meaningless single voxel activations. Anatomical locations were identified using the SPM Anatomy Toolbox 2.2b (Eickhoff et al. 2005) and the Harvard–Oxford cortical structural atlas (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>).

Wherever reported, percent signal change was extracted using the MarsBaR toolbox (version 0.44; <http://marsbar.sourceforge.net/>). To this end, we created 6 mm spheres around the peak voxel within the respective cluster in each contrast and extracted parameter estimates.

Psychophysiological interaction (PPI) analyses

Task-related functional connectivity between conditions of interest was assessed with a generalized psychophysiological interaction analysis (gPPI, McLaren, Ries, Xu, & Johnson 2012). Seed volumes of interest (VOI) were defined

by drawing 6 mm spheres around each subject's individual nearest activated voxel relative to the group peak of a given contrast at a threshold of $p < 0.05$. This lenient threshold ensured that each participant's VOI was in the same anatomical region as the group peak. To explore functional connectivity between key semantic regions and other brain areas we seeded from the following activation peaks: left aIFG, anterior AG (PGa), posterior AG (PGp), pMTG, ATL and DMPFC. The design matrix of each participant for each VOI comprised (1) the deconvolved time series of the first eigenvariate of the BOLD signal from the VOI, forming the physiological variable, (2) each condition convolved with the HRF, forming the psychological variable, and (3) the interaction of the psychological and physiological variable, forming the PPI term. At the single-subject level, whole-brain GLMs were conducted creating 3 contrasts (of the PPI terms) for each VOI model based on the univariate GLM results: 1) meaningful > anomalous, 2) meaningful > pseudowords, 3) anomalous > pseudowords. At the group level, we conducted one-sample t-tests for each contrast of interest. Contrast images were thresholded at $p < 0.05$, cluster-level family wise error (FWE) corrected, with a voxel-wise threshold of $p < 0.001$.

Results

Behavioral results

Overall accuracy was high in both tasks (mean implicit: 94.14%, mean explicit: 95.75%). In the implicit task (comparison of the lexical status of the words), meaningful phrases had a significantly higher accuracy than both anomalous and pseudoword phrases. In the explicit task (meaningfulness judgement of the phrase), an opposite picture emerged with pseudoword phrases being significantly more accurate than anomalous and meaningful phrases (Fig. 2a, c; SI Table 3). Reaction times in the implicit task were significantly faster for meaningful than anomalous and pseudoword phrases and for anomalous relative to pseudoword phrases. In the explicit task, reaction times were significantly faster for pseudoword phrases than meaningful and anomalous phrases and faster for meaningful relative to anomalous phrases (Fig. 2b, d; SI Table 3).

fMRI results

Explicit task: meaningful composition (meaningful > anomalous phrases)

To determine regions that guide meaningful composition, we compared meaningful versus anomalous phrases in the explicit task. Here, we found significantly increased activity

in the anterior part of left AG (PGa) extending into posterior supramarginal gyrus (SMG) and intraparietal sulcus (IPS), ACC, left pMTG, left ventromedial prefrontal cortex (vmPFC) and a small cluster in the posterior part of the right AG (PGp) (Fig. 3A; SI Table 4). As evident from the percent signal change for the left PGa ROI in SI Fig. 1A, we found positively increased task-related activity relative to rest for meaningful phrases and decreased activity below baseline for the other two conditions.

The opposite contrast of anomalous versus meaningful phrases did not yield any significant activation differences.

Explicit task: meaningful > pseudoword phrases

Comparing meaningful phrases with pseudoword phrases in the explicit task yielded activation in a wide-spread largely left-lateralized network of regions comprising aIFG (pars orbitalis), DMPFC, AG (PGp) extending into PGa, SMG and IPS, pMTG, pITG, ATL (including temporal pole), ACC (extending to right ACC) and posterior cingulate cortex (PCC), cerebellum (crus I/II), precuneus, insula and hippocampus. Additionally, increased right-hemispheric activity was found in the cerebellum, insula (extending into temporal pole), primary motor area (M1), fusiform gyrus, pITG and AG (PGp) (Fig. 3B; SI Table 5). In light of the discussion about the role of the left AG in the DMN and its potential *deactivation* during semantic tasks relative to rest (Humphreys et al. 2015), we additionally investigated the response pattern of the left PGp cluster. These analyses yielded deactivations for all three conditions compared to resting baseline (SI Fig. 1b).

Explicit task: anomalous > pseudoword phrases

A similar pattern emerged for the contrast of anomalous versus pseudoword phrases. This contrast yielded a largely left-lateralized network of regions including aIFG (pars orbitalis/triangularis), DMPFC, pMTG, ATL, AG (PGp), fusiform gyrus and thalamus. Right hemispheric activation comprised the cerebellum, aIFG/insula, amygdala and thalamus (Fig. 3c; SI Table 6).

Conjunction analysis: explicit meaningful > pseudoword phrases and anomalous > pseudoword phrases

To explore regions that are activated independently of the meaningfulness of the final phrase, we conducted a conjunction analysis of the contrasts meaningful > pseudowords \cap anomalous > pseudowords. This conjunction will be referred to as *general phrasal composition*, to emphasize that it does not depend on the meaningfulness. This yielded common activations for all regions that were also involved in anomalous > pseudoword processing, showing

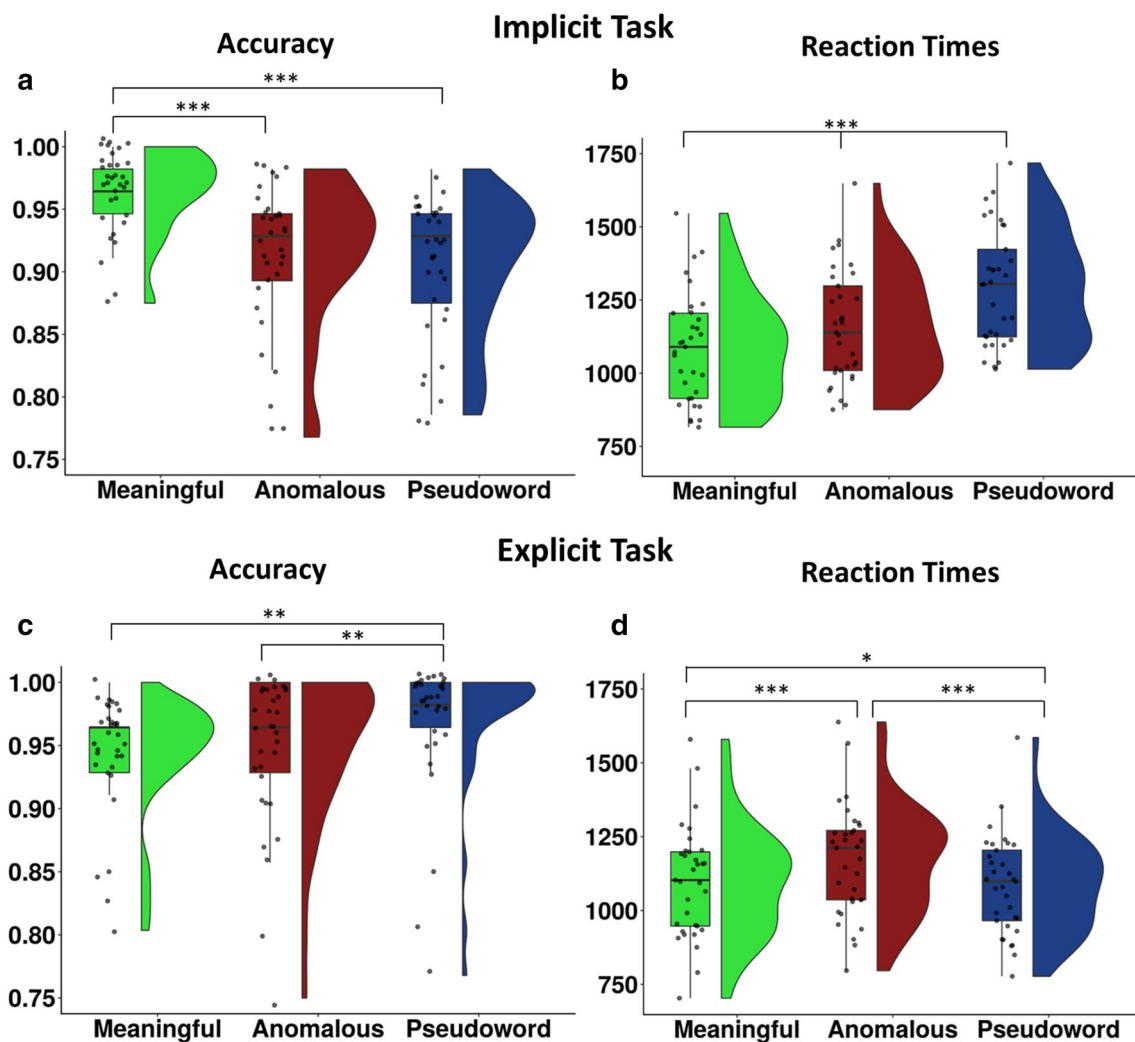


Fig. 2 Raincloud plots (Allen et al. 2019) illustrating the data distribution of each participant's mean behavioral scores and boxplots overlaid with individual mean data points for the implicit (a, b) and explicit task (c, d). * $p < .05$; ** $p < 0.01$; *** $p < 0.001$

that *this contrast* engages a subset of the regions for *meaningful > pseudoword phrases* (Fig. 3d; SI Table 7). Thus, the left aIFG, ATL, pMTG, PGp, DMPFC, thalamus, right aIFG and cerebellum appear to be involved in phrasal composition independently of the plausibility of the resulting phrase.

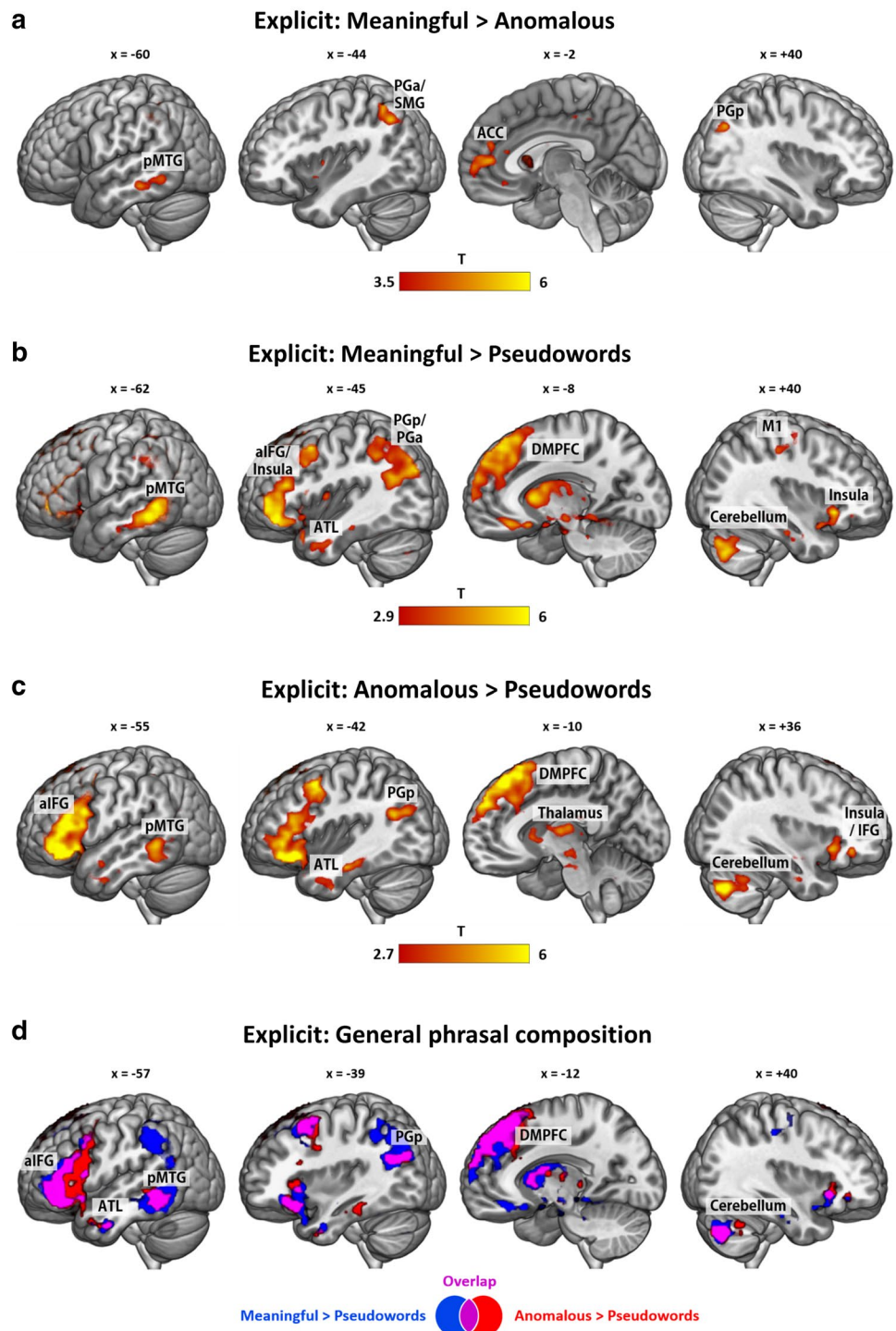
Implicit task: meaningful > pseudoword phrases

In the implicit task, only the contrast of meaningful > pseudoword phrases yielded significant results. Here, we found increased activation in left AG (PGp), dorsomedial prefrontal cortex (DMPFC), pMTG/ITG and a small cluster in the right anterior cingulate cortex (ACC) (Fig. 4; SI Table 8).

Task-independent activation for meaningful > pseudoword phrases

After exploring the contrasts for each task separately, we were interested in common regions that are activated independently of the task. To this end, we conducted conjunction analyses of the only contrast that yielded significant results in both tasks: explicit *meaningful > pseudoword phrases* \cap implicit *meaningful > pseudoword phrases*. We observed a significant cluster in the left PGp and very small clusters in the left pITG and DMPFC (Fig. 5; SI Table 9). Thus, the only regions showing task-independent activations for *meaningful > pseudoword phrases* are the posterior part of the AG and to a lesser extent, parts of DMPFC and pITG.

Fig. 3 Whole-brain activations in the explicit task for the contrast **a** meaningful > anomalous, **b** meaningful > pseudowords, **c** anomalous > pseudowords and **d** overlap (purple) of the contrasts meaningful > pseudowords (blue) and anomalous > pseudowords (red), resulting in *general phrasal composition*. All activation maps are thresholded at $q < 0.05$ FDR-corrected. ACC=anterior cingulate cortex, aIFG=anterior inferior frontal gyrus, ATL=anterior temporal lobe, DMPFC=dorsomedial prefrontal cortex, pMTG=posterior middle temporal gyrus, PGa=angular gyrus anterior division, PGp=angular gyrus posterior division, SMG=supramarginal gyrus



Task-specific activations

From the results above, it appears that brain regions exist which are selectively involved during the explicit task, but not during the implicit one. To further substantiate this finding, we conducted paired t-tests for all contrasts between the two tasks.

For the contrast of *meaningful > pseudoword phrases*, paired t-tests confirmed that left aIFG (extending into insula), DMPFC, pMTG/ITG, IPS/SMG and right aIFG, thalamus and cerebellum are significantly more involved in the explicit than in the implicit task (Fig. 6A; SI Table 10).

The contrast *anomalous > pseudoword phrases* also revealed task-dependent activations in left aIFG, DMPFC, right cerebellum, aIFG and thalamus (Fig. 6B,

Fig. 4 Whole-brain activations in the implicit task for the contrast meaningful > pseudoword phrases, thresholded at $q < 0.05$ FDR-corrected. DMPFC = dorsomedial prefrontal cortex, pMTG = posterior middle temporal gyrus, pITG = posterior inferior temporal gyrus, PGp = angular gyrus posterior division

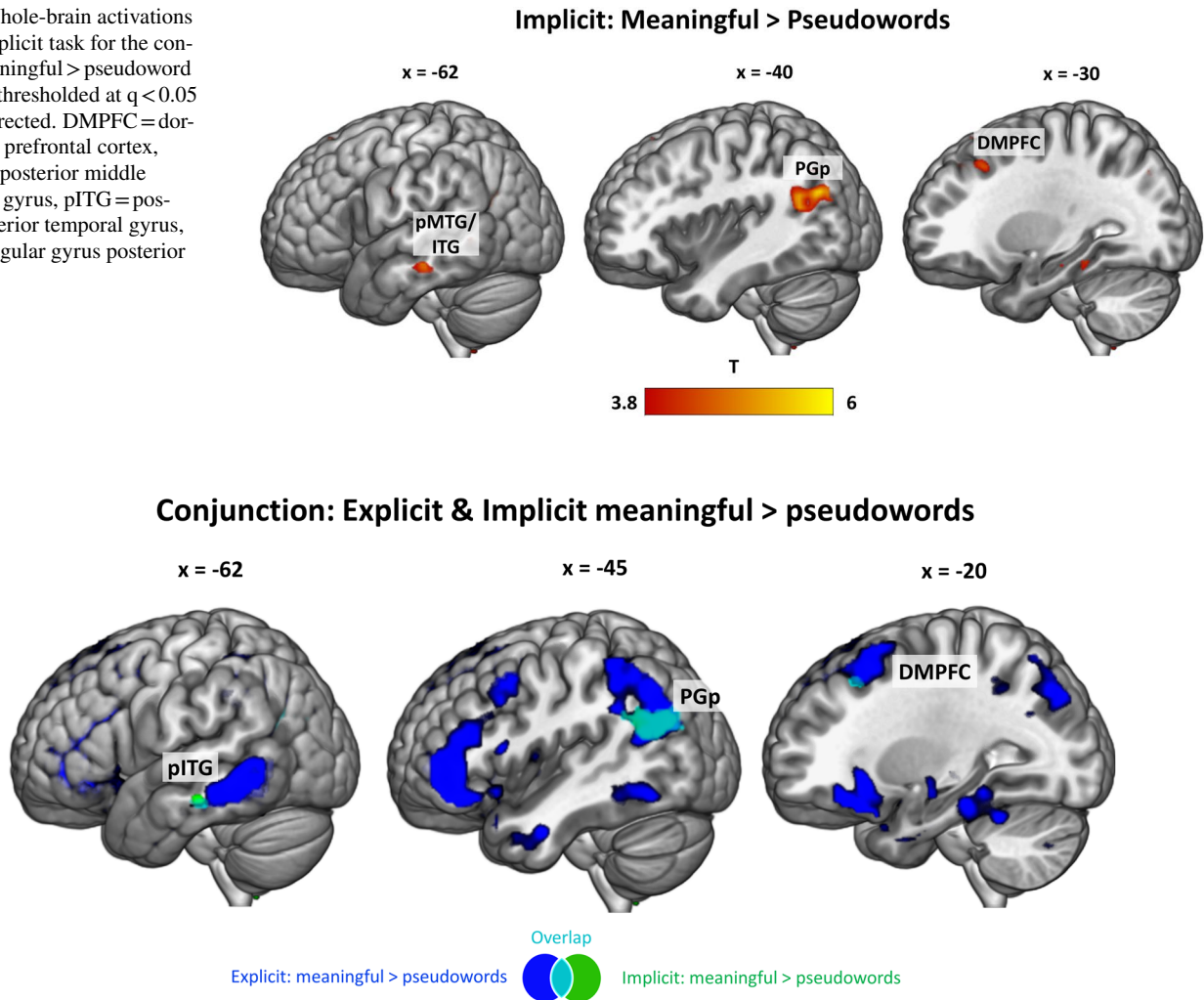


Fig. 5 Task-independent activations: Overlap (cyan) between the contrast meaningful > pseudowords in the explicit task (blue) and the implicit task (green), thresholded at $q < 0.05$ FDR-corrected.

DMPFC = dorsomedial prefrontal cortex, pITG = posterior inferior temporal gyrus, PGp = angular gyrus posterior division

SI Table 11). Thus, these regions seem to be selectively involved when explicit meaningfulness judgement is required.

For the more specific contrast *meaningful > anomalous phrases*, this only yielded significant activations when lowering the threshold to cluster-level FWE correction ($p < 0.05$) in PGa and ACC. Thus, we can only cautiously speak of a trend of task dependence in these regions.

Psychophysiological interaction (PPI) analysis for meaningful > pseudoword phrases

Finally, we set up several PPI models to investigate task-specific interactions. In the explicit task, we found significant functional coupling between the left PGp (as seed region) and left aIFG (pars triangularis; BA45) and bilateral

pre-supplementary motor cortex for meaningful > pseudowords; Fig. 7; SI Table 12). No other seed region or contrast yielded significant results.

Discussion

In the present study, we sought to characterize the neural correlates of basic semantic composition and the task-dependent network interactions characterizing such process. Despite numerous efforts to identify the key regions of combinatorial semantic processing in previous neuroimaging studies, several questions remained open.

Here, we developed a paradigm that is sensitive to capture both the specific process of combining two words into a meaningful phrase (meaningful > anomalous) as well as the more general process of accessing two words versus only

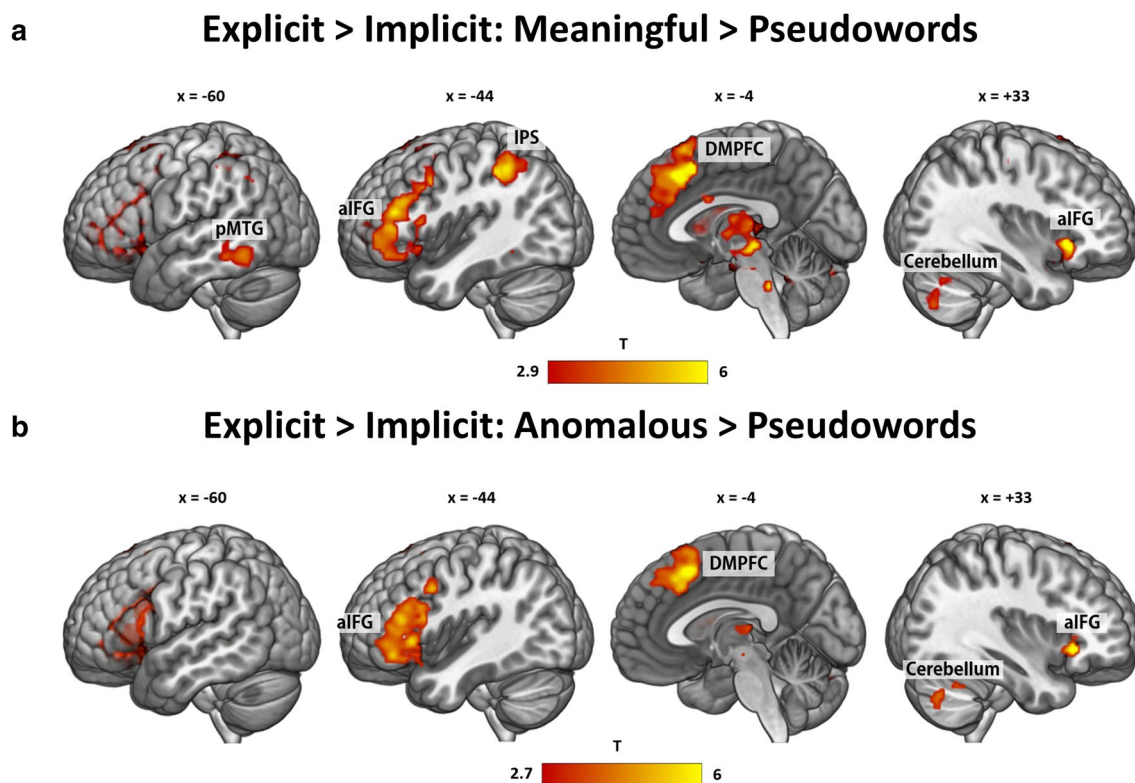


Fig. 6 Task-dependent activations: **a** Stronger activation during the explicit than the implicit task for meaningful > pseudowords phrases. **b** Stronger activation during the explicit than the implicit task for anomalous > pseudoword phrases. Thresholded at $q < 0.05$ FDR-cor-

rected. aIFG = anterior inferior frontal gyrus, DMPFC = dorsomedial prefrontal cortex, IPS = inferior parietal sulcus, pMTG = posterior middle temporal gyrus

one word, comprising the contrasts meaningful > pseudoword and anomalous > pseudoword (together referred to as *general phrasal composition*). We identified separable neural correlates for these two processes. The more specific *meaningful composition* (meaningful > anomalous) engages a small set of mostly left-hemispheric regions (anterior AG/posterior SMG, pMTG, ACC), which seem to lie at the heart of specific meaning composition, necessary for evaluating the plausibility of a phrase.

In contrast, the more *general phrasal composition* process, identified by the conjunction of meaningful > pseudoword and anomalous > pseudoword phrases, appears to be largely independent of the plausibility of the resulting phrase. This process is associated with a widely distributed pattern of left-lateralized activity, including the aIFG, ATL, DMPFC, pMTG and AG. Crucially, the posterior part of the AG is involved in a task-independent manner, suggesting a role in general semantic representation processes that goes beyond task-specific activation. This region might thus reflect automatic semantic processing. In contrast, aIFG, a large part of DMPFC, pMTG and IPS/SMG show stronger

engagement for the explicit task in the contrast meaningful > pseudoword phrases, than in the supposedly more lexical-level implicit task. This process furthermore relies on the functional interaction between the left posterior AG, the aIFG and the pre-supplementary motor area, as shown in our PPI analysis.

While previous research has proposed aIFG, ATL and AG to play a key role in semantic composition, our results help to disentangle the subprocesses that guide semantic composition in these regions. Our results may suggest that semantic composition consists of at least two processes. The first process requires the combination of retrieved semantic information of each constituent, which we here termed *phrasal composition*. The second process reflects the plausibility evaluation of the combined concept based on prior knowledge. Whereas both of our two real-word conditions (meaningful and anomalous) result in success of the first process, only meaningful phrases also result in plausible composition during the second process—the *meaningful composition*. Note that we cannot distinguish whether both processes take place in parallel or after each other since the temporal resolution of our fMRI design is too low to allow for disentangling these processes.

Task-specific functional coupling for meaningful > pseudowords

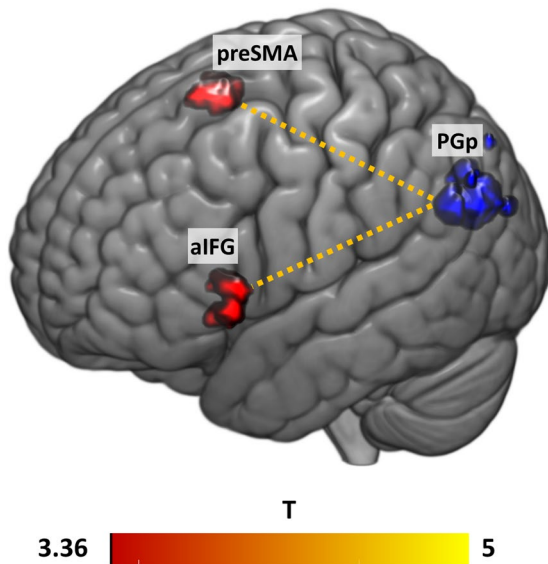


Fig. 7 gPPI results with PGp as seed region (blue; 6 mm spheres around individual peak activation) for the contrast meaningful > pseudowords in the explicit task (thresholded at $p < 0.05$ FWE-corrected at the cluster level). For visualization of the individual seed regions, we have plotted the individual 6 mm spheres from each participant on top of each other

Dissociating the timing was not the main goal of our study but it is plausible to assume that the first process (*phrasal composition*) occurs within the time frame of the typically observed ATL effect from MEG studies at around 200–250 ms (Pylkkänen 2019) whereas the plausibility evaluation occurs at a later time point, possibly around the typical N400 effect (Kutas and Federmeier 2011). This distinction into several subprocesses is also in line with current neurocognitive models of conceptual combination (Coutanche et al. 2019; Westerlund and Pylkkänen 2017).

Meaningful composition (meaningful > anomalous) engages the anterior AG and semantic control regions

In our first contrast of meaningful versus anomalous phrases, we found increased activity in several semantic regions, mainly in the left hemisphere. These regions included the left anterior AG (PGa) and neighboring SMG / IPS, the left pMTG, the bilateral ACC and the left vmPFC, as well as several smaller clusters in the right hemisphere, including the right posterior AG (PGp).

The strong contribution of the PGa is well in line with a previous study that used a similar task (Price et al. 2015). In that study, AG (with a peak in PGa) showed increased

activity for more meaningful relative to less meaningful adjective-noun phrases. Recently, left AG was found for the processing of verb phrases and noun phrases relative to word lists (Matchin et al. 2019a, b), providing further evidence for its key role in the meaningful conceptual composition. The functional relevance of the left AG for semantic composition was demonstrated in lesion and neurostimulation studies (Price et al. 2015, 2016). Regarding the role of different AG subregions in semantic processing, previous meta-analyses have identified a dorsal, middle and ventral subdivision, each serving a different function with respect to semantic tasks (Noonan et al. 2013; Seghier 2013). Crucially, our activation cluster for *meaningful composition* was located at the border of PGa, posterior SMG (PFm) and IPS. This overlaps with the dorsal AG subregion identified by Noonan et al. (2013) that was associated with semantic control and conceptual combination. Additionally, the analysis of percent signal change revealed a task-positive engagement of PGa for meaningful phrases. While this could be interpreted as reflecting part of the multiple demand network, the condition that elicited more activation was indeed the easier (faster) one (meaningful versus anomalous). Thus, we propose that PGa likely reflects successful semantic plausibility evaluation.

Aside from the left AG, our results also revealed a small cluster in the right AG for *meaningful composition*. This activation is in line with two previous studies that reported right AG involvement in basic combinatorial processing (Graves et al. 2010; Price et al. 2015). Graves and colleagues proposed that while left temporo-parietal regions represent single word meaning, the right hemisphere represents the overlap of single concepts and combines the meaning of them, which might explain the observed upregulation of the right AG in the present study.

Another semantic region that we found was the left pMTG. This finding was a bit surprising, as the pMTG has not classically been ascribed a key role in semantic composition. Rather, converging evidence from neuroimaging and neurostimulation studies has attributed the pMTG (together with the anterior IFG) a role in semantic control during semantic association tasks at the word level (Davey et al. 2016; Noonan et al. 2013; Whitney et al. 2011). Our results fit with this view insofar as we found much stronger activation in pMTG in the explicit than the implicit task. It is reasonable to assume that the implicit task does not require the same level of executive semantic processing as the explicit task. The observed engagement of the pMTG in all real-word contrasts (i.e., in explicit *meaningful > anomalous*, *meaningful > pseudowords* and *anomalous > pseudowords*) might thus indicate that this region generally increases its activation towards more meaningful versus less meaningful stimuli.

Interestingly, we also found engagement of bilateral ACC for *meaningful composition*. This region is typically involved in error monitoring tasks (Botvinick et al. 2004), a process that is unlikely to occur during meaningful composition. It is furthermore a key region in the cingulo-opercular control network and associated with task maintenance (Vaden et al. 2013). However, there is recent evidence that ACC is also involved in semantic tasks. Almeida and colleagues (2016) found widespread activations including ACC for indeterminate sentences (*The author began the book*) as compared to preferred sentences (*The author wrote the book*) and anomalous sentences (*The author drank the book*). They interpret the role of these regions as employing pragmatic-inferential processes. Furthermore, the meta-analysis by Noonan et al. (2013) identified the ACC as part of the wide-spread semantic control network. The fact that we found ACC involvement only in the explicit task supports this view.

Finally, we also observed a small cluster in the left vmPFC for *meaningful composition*. This region has been found in a number of MEG studies and has recently been associated with representing the final combinatory output (Pylkkänen 2019), which would fit with our findings.

Notably, we did not find significant task-related activity during the implicit task for this contrast. The lack of activation differences in the implicit task could suggest that meaning is not automatically composed but might rather be restricted to situations where it is task-relevant. However, the observed significant differences in accuracy and response times between meaningful and anomalous phrases in the implicit task might indicate that participants did automatically evaluate the meaning of the phrases. Crucially, the lexical material of meaningful and anomalous phrases did not differ, so participants should be equally successful in deciding whether the stimuli are both real words or not. We thus believe that the lower accuracy for the anomalous relative to the meaningful condition in the implicit task reflects an automatic plausibility judgement. A possible explanation for why we did not find activation differences in the fMRI results could be that the nature of our lexical status task required increased task demands, which were similar for meaningful and anomalous phrases. Overall, we believe that our explicit task required deeper semantic processing while the implicit task showed a higher degree of domain-general cognitive demands and thus automatic semantic composition was only observable at the behavioral and not at the neural level. An intriguing question for future studies would be whether other implicit semantic tasks (e.g. classical lexical decision, phonological tasks) are able to detect combinatorial processing differences in fMRI data for meaningful versus anomalous phrases, which, in the current study, are only visible at the behavioral level.

Additionally, it should be noted that we did not observe significant activation differences for the reversed contrast

in the explicit task (i.e., for *anomalous* relative to *meaningful* phrases). While it is conceivable to find a similar effect in the aIFG as reported for ambiguous > unambiguous sentences (e.g. Rodd et al. 2005) or the classical N400 in the EEG literature, previous fMRI studies have also found that the latter effect is a lot weaker in the hemodynamic modality (Lau and Namyst 2019). This might be due to the much lower time-sensitivity of fMRI compared to EEG, with the latter thus being better suited to capture the relatively short-lasting N400 effect. Another possible explanation is the nature of our task: Having to judge the meaningfulness might imply that there would be meaning in the phrases. Participants actively searched for associations and meaning to make their judgement and were thus already primed towards the more meaningful condition (Kuperberg, 2007).

General phrasal composition engages a large semantic network in the left hemisphere

In line with previous research, we found recruitment of left AG, aIFG and ATL, along with other classical regions of the semantic network (e.g., left pMTG and DMPFC) and right-hemispheric motor (control) regions (cerebellum, premotor, primary and somatosensory motor cortex) for the contrast of meaningful versus pseudoword phrases in the explicit task. Interestingly, our results suggest that the plausibility of the final concept with respect to existing world knowledge does not seem to be relevant for the contribution of the core semantic network, as we observed large overlap of the two contrasts *meaningful > pseudowords* and *anomalous > pseudowords*. The identified left-hemispheric language network encompasses brain areas that are sensitive to the amount of semantic information that can be integrated. This finding nicely fits with a recent proposal that the core process (“driver”) of the language network is composition (Mollica et al. 2020).

Consistent with its proposed role in executive semantic control (Chiou et al. 2018; Lambon Ralph et al. 2016; Noonan et al. 2013; Whitney et al. 2011), the left aIFG was selectively activated in the explicit task, for both real-word conditions compared to the pseudoword condition. Previous studies have associated the left aIFG with monitoring and selecting semantic information when several alternatives are present (Binder and Desai 2011; Lau et al. 2008; Noonan et al. 2013; Whitney et al. 2012). In our study, the aIFG activates as a function of the amount of words that can be retrieved and combined, independent of the meaningfulness of the final phrase as it was observed both for the contrast of *meaningful > pseudoword* and *anomalous > pseudoword phrases*. This fits with the results from Schell et al. (2017), who found aIFG activation for adjective-noun phrases (*blue boat*) versus single words (*boat*). Consequently, aIFG involvement might reflect increased semantic load carried

by real words, which also goes beyond the sentential level (Zaccarella et al. 2015).

The ATL has previously been suggested to be the key conceptual composition hub (e.g. Baron et al. 2011; Bemis and Pykkänen 2011; Westerlund and Pykkänen 2017). In this study we found evidence for its contribution only at the more general phrasal level in the explicit task (i.e., for the contrasts meaningful > pseudowords and anomalous > pseudowords) and not for the more specific *meaningful composition* (*meaningful* > *anomalous*). While previous studies have not aimed to differentiate between the specific processes during semantic composition, our distinction into a phrasal composition and a plausibility evaluation step provides an explanation for why some studies did not find ATL involvement for basic semantic composition tasks. The ATL does not seem to be relevant for the evaluation of plausibility but rather aids conceptual combination at the phrasal level. Additionally, our results do not lend support to the notion that the ATL guides the automatic composition of concepts (Bemis and Pykkänen 2013a), as we only find involvement of this region in the explicit task. For further discussion of the role of the ATL in conceptual combination, please also see a recent study by Kochari et al. (2020).

Our study additionally revealed a strong engagement of DMPFC during explicit *general phrasal composition*. This converges with a previous study by Graves et al. (2010), who also found DMPFC engagement selectively for explicit meaningful judgments. Accordingly, Binder and colleagues (2009) identified the DMPFC as a core semantic region. Together with the left IFG, the DMPFC was proposed to act as a control region that guides goal-directed retrieval of conceptual information stored in temporo-parietal cortices (Binder and Desai 2011). Consequently, DMPFC might be a semantic control region that has been overlooked in most studies. Indeed, previous reviews have failed to acknowledge its role in semantic tasks and a consistent theory of its exact contribution to semantic processing is still lacking. Note, however, that some studies found DMPFC activation in tasks with high control demands and consequently ascribed it to the multiple demand network (Assem et al. 2019; Geranmayeh et al. 2017; Noonan et al. 2013). Since we found stronger DMPFC involvement in the conditions that required longer reaction times, this could be an alternative explanation for the observed DMPFC activation. Overall, our study cannot distinguish between the two alternative interpretations of DMPFC activation reflecting specific semantic vs. domain-general control demands.

Regarding the subdivision of the AG and the specific roles of its subregions in semantic processing, our activation cluster for *the conjunction of meaningful > pseudoword and anomalous > pseudoword phrases* in the posterior division PGp overlaps with the mid-AG region identified in previous work (Noonan et al. 2013: MNI coordinates

x, y, z: -39, -69, 30; Seghier, 2010: MNI coordinates x, y, z: -48, -68, 29). Despite partly contradictory findings, this subregion was consistently reported for processing concrete relative to abstract concepts and was thus ascribed a role in semantic representation for rich multimodal concepts. As we found relatively stronger involvement of PGp in both real word conditions as compared to the pseudoword condition, our results fit with the account that this subregion codes for the semantic richness. More conceptual information can be retrieved during the two real-word conditions than for the pseudoword condition, but this distinction does not hold when the two real-word conditions are compared directly. Thus, we did not observe engagement of left PGp in the specific contrast *meaningful > anomalous phrases* but only in the conjunction of the two contrasts versus pseudoword phrases. We believe that our results help to specify the role of different subregions in the left AG in semantic composition. The posterior cluster activates as a function of the amount of lexical information that can be integrated, while the anterior part seems to be relevant for the plausibility evaluation of the phrase. Note further that PGp, together with small clusters in the pITG and DMPFC, were the only regions that were also involved in the implicit task, speaking for a task-independence of these regions in lexical-semantic processes. Thus, in our view, PGp is involved in the representation of more versus less conceptual information regardless of the plausibility of the resulting phrase and independent of the task. Conversely, a region in the parietal lobe that shows strong task-dependent involvement is the IPS, bordering posterior SMG. This region was significantly more involved during the explicit than the implicit task and thus likely guides controlled conceptual retrieval together with the aIFG and DMPFC.

Interestingly, analyses of percent signal change revealed distinct patterns in the two AG subregions: While PGa shows positive activation only for meaningful phrases and not for anomalous or pseudoword phrases, PGp shows relatively less deactivation for meaningful and anomalous than pseudoword phrases (SI Fig. 1). Importantly, the observed relative engagement of PGp cannot be explained in terms of lower task difficulty, as we found the opposite effect: Meaningful and anomalous phrases were processed less accurately and slower than pseudoword phrases in the explicit task but involved PGp relatively stronger. Consequently, PGp does not show the classical pattern of a default mode region (less deactivation for easy versus hard task, cf. Humphreys et al. 2015; Humphreys and Lambon Ralph 2015) and fits with the account of a key semantic region.

Fronto-parietal interactions for meaningful > pseudoword phrases

Aside from task-related activations, our results revealed that during the processing of meaningful phrases, PGp shows increased functional connectivity with aIFG and pre-SMA. The task-related functional coupling was stronger for the processing of meaningful phrases compared to pseudoword phrases. These findings extend previous results of functional connectivity at the single word level. In a TMS study, Hartwigsen et al. (2015) showed that temporary disruption of either aIFG or AG alone did not lead to a significant impairment of semantic decisions, while combined TMS over both regions significantly delayed reaction times in the semantic task. This suggests that AG and aIFG can compensate for the disruption of the respective other node. Moreover, after longer-lasting disruption of left AG, this area had an inhibitory influence on the left aIFG during semantic word decisions, further substantiating the strong interaction between both regions (Hartwigsen et al. 2017). Crucially, our results provide novel supporting evidence for the notion of a semantic network involving AG and aIFG not only at the single word but also at the basic combinatorial level. Aside from aIFG, task-related functional coupling was also increased between PGp and the pre-SMA. While the pre-SMA does not belong to the classical language network, there is evidence that it is involved in higher-order cognitive processes such as semantic processing, independent of motor effects (Hertrich et al. 2016). In summary, the results from our functional connectivity analysis significantly extend previous findings of changes in task-related activity during semantic composition and provide new insight into functional interactions during meaningful phrasal compositions at the network level. We note that one limitation of our PPI analyses is that from the 6 seed regions, only the PGp yielded significant task-specific changes in connectivity with any other brain region.

Conclusion

In the present study, we identified distinct neural signatures for two processes during explicit basic semantic composition: A *meaningful composition* process, strongly dependent on the resulting plausibility and a *general phrasal composition* process, which is independent of the plausibility of the resulting phrase. The latter process engages a widespread semantic network in the left hemisphere, including PGp, aIFG, DMPFC and large parts of the temporal lobe. Crucially, only PGp shows task-independent engagement, pointing towards a role in automatic semantic processing. PGp furthermore strongly interacts with the pre-SMA and another key semantic region in the left hemisphere, the

aIFG, thus forming the core semantic network of phrasal composition. The more specific *meaningful composition* engages a subset of the semantic control regions found for phrasal composition, and the left anterior AG, bordering posterior SMG. Consequently, AG appears to be decomposable into distinct subregions during semantic composition: PGp codes for the amount of conceptual information that can be integrated, while PGa (together with PFm) evaluates phrasal plausibility.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00429-020-02196-2>.

Acknowledgements We wish to thank Lisa Kunz for help during data acquisition and Nicole Pampus for acquiring participants. We also like to thank Laura Nieberlein for her assistance in piloting the experiment. Moreover, we thank Angela D. Friederici for helpful discussions and Philipp Kuhnke and Anna Rysop for comments on this manuscript. This work was supported by the Max Planck Society. GH is supported by the German Research Foundation (DFG, HA 6314/3-1, HA 6314/4-1). We thank the University of Minnesota Center for Magnetic Resonance Research for the provision of the multiband EPI sequence software.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Astrid Graessner, Emiliano Zaccarella and Gesa Hartwigsen. The first draft of the manuscript was written by Astrid Graessner and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Gesa Hartwigsen is supported by the German Research Foundation (DFG, HA 6314/3–1, HA 6314/4–1).

Availability of data and material The data that support the findings of this study are available on request from the corresponding author.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

Ethical approval The study protocol conformed to the principles of the Declaration of Helsinki and was approved by the local ethics committee at the University of Leipzig. (Ethics approval number: 155/17-ek).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA (2019) Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res* 4:63
- Assem M, Blank IA, Mineroff Z, Ademoglu A, Fedorenko E (2019) Activity in the Fronto-Parietal Multiple-demand Network is Robustly Associated with Individual Differences in Working Memory and Fluid Intelligence. *bioRxiv* 110270.
- Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12(5):193–200. <https://doi.org/10.1016/j.tics.2008.02.004>
- Baron SG, Osherson D (2011) Evidence for conceptual combination in the left anterior temporal lobe. *Neuroimage* 55(4):1847–1852. <https://doi.org/10.1016/j.neuroimage.2011.01.066>
- Bassett DS, Sporns O (2017) Network neuroscience. *Nat Neurosci* 20(3):353–364. <https://doi.org/10.1038/nn.4502>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- Bemis DK, Pykkänen L (2011) Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J Neurosci* 31(8):2801–2814. <https://doi.org/10.1523/JNEUROSCI.5003-10.2011>
- Bemis DK, Pykkänen L (2013a) Flexible composition: MEG evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS ONE* 8(9):e73949. <https://doi.org/10.1371/journal.pone.0073949>
- Bemis DK, Pykkänen L (2013b) Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb Cortex* 23(8):1859–1873. <https://doi.org/10.1093/cercor/bhs170>
- Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends Cogn Sci* 15(11):527–536. <https://doi.org/10.1016/j.tics.2011.10.001>
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19(12):2767–2796. <https://doi.org/10.1093/cercor/bhp055>
- Blanco-Elorrieta E, Kastner I, Emmorey K, Pykkänen L (2018) Shared neural correlates for building phrases in signed and spoken language. *Sci Rep* 8(1):5492. <https://doi.org/10.1038/s41598-018-23915-0>
- Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci* 8(12):539–546
- Boylan C, Trueswell JC, Thompson-Schill SL (2015) Compositionality and the angular gyrus: a multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia* 78:130–141
- Boylan C, Trueswell JC, Thompson-Schill SL (2017) Relational vs. attributive interpretation of nominal compounds differentially engages angular gyrus and anterior temporal lobe. *Brain Lang* 169:8–21
- Brennan J, Pykkänen L (2012) The time-course and spatial distribution of brain activity associated with sentence processing. *Neuroimage* 60(2):1139–1148
- Brysbaert M, Buchmeier M, Conrad M, Jacobs AM, Bolte J, Bohl A (2011) The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp Psychol* 58(5):412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Brysbaert M, Warriner AB, Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. *Behav Res Methods* 46(3):904–911
- Chiou R, Humphreys GF, Jung J, Lambon Ralph MA (2018) Controlled semantic cognition relies upon dynamic and flexible interactions between the executive "semantic control" and hub-and-spoke "semantic representation" systems. *Cortex* 103:100–116. <https://doi.org/10.1016/j.cortex.2018.02.018>
- Coutanche M, Solomon S, Thompson-Schill S (2019) Conceptual combination. *The cognitive neurosciences*.
- Davey J, Thompson HE, Hallam G, Karapanagiotidis T, Murphy C, De Caso I, Krieger-Redwood K, Bernhardt BC, Smallwood J, Jefferies E (2016) Exploring the role of the posterior middle temporal gyrus in semantic cognition: integration of anterior temporal lobe with executive processes. *Neuroimage* 137:165–177. <https://doi.org/10.1016/j.neuroimage.2016.05.051>
- de Almeida RG, Riven L, Manouilidou C, Lungu O, Dwivedi VD, Jarema G, Gillon B (2016) The neuronal correlates of indeterminate sentence comprehension: an fMRI study. *Front Hum Neurosci* 10:614. <https://doi.org/10.3389/fnhum.2016.00614>
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25(4):1325–1335
- Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, Miller KL, Ugarbil K, Yacoub E (2010) Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS ONE* 5(12):e15710
- Geranmayeh F, Chau TW, Wise RJ, Leech R, Hampshire A (2017) Domain-general subregions of the medial prefrontal cortex contribute to recovery of language after stroke. *Brain* 140(7):1947–1958
- Graves WW, Binder JR, Desai RH, Conant LL, Seidenberg MS (2010) Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage* 53(2):638–646. <https://doi.org/10.1016/j.neuroimage.2010.06.055>
- Halai AD, Welbourne SR, Embleton K, Parkes LM (2014) A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Hum Brain Mapp* 35(8):4118–4128. <https://doi.org/10.1002/hbm.22463>
- Hardwick RM, Caspers S, Eickhoff SB, Swinnen SP (2018) Neural correlates of action: comparing meta-analyses of imagery, observation, and execution. *Neurosci Biobehav Rev* 94:31–44
- Hartwigsen G (2018) Flexible redistribution in cognitive networks. *Trends Cogn Sci* 22(8):687–698. <https://doi.org/10.1016/j.tics.2018.05.008>
- Hartwigsen G, Bzdok D, Klein M, Wawrzyniak M, Stockert A, Wrede K, Classen J, Saur D (2017) Rapid short-term reorganization in the language network. *Elife*. <https://doi.org/10.7554/eLife.25964>
- Hartwigsen G, Weigel A, Schuschan P, Siebner HR, Weise D, Classen J, Saur D (2015) Dissociating parieto-frontal networks for phonological and semantic word decisions: a condition-and-perturb TMS study. *Cereb Cortex* 26(6):2590–2601. <https://doi.org/10.1093/cercor/bhv092>
- Hertrich I, Dietrich S, Ackermann H (2016) The role of the supplementary motor area for speech and language processing. *Neurosci Biobehav Rev* 68:602–610
- Humphreys GF, Hoffman P, Visser M, Binney RJ, Lambon Ralph MA (2015) Establishing task- and modality-dependent dissociations between the semantic and default mode networks. *Proc Natl Acad Sci USA* 112(25):7857–7862. <https://doi.org/10.1073/pnas.1422760112>
- Humphreys GF, Lambon Ralph MA (2015) Fusion and fission of cognitive functions in the human parietal cortex. *Cereb Cortex* 25(10):3547–3560
- Humphries C, Binder JR, Medler DA, Liebenthal E (2006) Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J Cogn Neurosci* 18(4):665–679

- Kanske P, Kotz SA (2010) Leipzig affective norms for German: a reliability study. *Behav Res Methods* 42(4):987–991. <https://doi.org/10.3758/BRM.42.4.987>
- Keuleers E, Brysbaert M (2010) Wuggy: a multilingual pseudoword generator. *Behav Res Methods* 42(3):627–633. <https://doi.org/10.3758/BRM.42.3.627>
- Kochari A, Lewis A, Schoffelen J-M, Schriefers H (2020) Semantic and syntactic composition of minimal adjective-noun phrases in Dutch: An MEG study. *bioRxiv*.
- Kuperberg GR (2007) Neural mechanisms of language comprehension: challenges to syntax. *Brain Res* 1146:23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>
- Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62:621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT (2016) The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18(1):42
- Lau EF, Namyst A (2019) fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity. *Brain Lang* 199:104697
- Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (De)constructing the N400. *Nat Rev Neurosci* 9(12):920–933. <https://doi.org/10.1038/nrn2532>
- Lewis GA, Poeppel D, Murphy GL (2018) Contrasting semantic versus inhibitory processing in the *Angular Gyrus*: an fMRI study. *Cereb Cortex*. <https://doi.org/10.1093/cercor/bhy118>
- Makuuchi M, Friederici AD (2013) Hierarchical functional connectivity between the core language system and the working memory system. *Cortex* 49(9):2416–2423. <https://doi.org/10.1016/j.cortex.2013.01.007>
- Matchin W, Brodbeck C, Hammerly C, Lau E (2019a) The temporal dynamics of structure and content in sentence comprehension: evidence from fMRI-constrained MEG. *Hum Brain Mapp* 40(2):663–678. <https://doi.org/10.1002/hbm.24403>
- Matchin W, Hammerly C, Lau E (2017) The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. *Cortex* 88:106–123. <https://doi.org/10.1016/j.cortex.2016.12.010>
- Matchin W, Liao CH, Gaston P, Lau E (2019b) Same words, different structures: an fMRI investigation of argument relations and the angular gyrus. *Neuropsychologia* 125:116–128. <https://doi.org/10.1016/j.neuropsychologia.2019.01.019>
- McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. *Neuroimage* 61(4):1277–1286
- Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Ugurbil K (2010) Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* 63(5):1144–1153
- Molinaro N, Paz-Alonso PM, Dunabeitia JA, Carreiras M (2015) Combinatorial semantics strengthens angular-anterior temporal coupling. *Cortex* 65:113–127. <https://doi.org/10.1016/j.cortex.2015.01.004>
- Mollica F, Siegelman M, Diachek E, Piantadosi ST, Mineroff Z, Futrell R, Kean H, Qian P, Fedorenko E (2020) Composition is the core driver of the language-selective network. *Neurobiology of Language* 1(1):104–134
- Mummery CJ, Patterson K, Price CJ, Ashburner J, Frackowiak RSJ, Hodges JR (2000) A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann Neurol* 47(1):36–45
- Nichols T, Brett M, Andersson J, Wager T, Poline J-B (2005) Valid conjunction inference with the minimum statistic. *Neuroimage* 25(3):653–660
- Noonan KA, Jefferies E, Visser M, Lambon Ralph MA (2013) Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal *Angular gyrus* and posterior middle temporal cortex. *J Cogn Neurosci* 25(11):1824–1850. https://doi.org/10.1162/jocn_a_00442
- Pallier C, Devauchelle AD, Dehaene S (2011) Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci* 108(6):2522–2527
- Price AR, Bonner MF, Peelle JE, Grossman M (2015) Converging evidence for the neuroanatomic basis of combinatorial semantics in the *Angular gyrus*. *J Neurosci* 35(7):3276–3284. <https://doi.org/10.1523/JNEUROSCI.3446-14.2015>
- Price AR, Peelle JE, Bonner MF, Grossman M, Hamilton RH (2016) Causal evidence for a mechanism of semantic integration in the angular gyrus as revealed by high-definition transcranial direct current stimulation. *J Neurosci* 36(13):3829–3838. <https://doi.org/10.1523/JNEUROSCI.3120-15.2016>
- Pylkkänen L (2019) The neural basis of combinatory syntax and semantics. *Science* 366(6461):62–66
- Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15(8):1261–1269
- Schell M, Zaccarella E, Friederici AD (2017) Differential cortical contribution of syntax and semantics: an fMRI study on two-word phrasal processing. *Cortex*. <https://doi.org/10.1016/j.cortex.2017.09.002>
- Seghier ML (2013) The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19(1):43–61. <https://doi.org/10.1177/1073858412440596>
- Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The cingulo-opercular network provides word-recognition benefit. *J Neurosci* 33(48):18979–18986
- Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. *J Cogn Neurosci* 14(4):550–560
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006) Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30(4):1414–1432. <https://doi.org/10.1016/j.neuroimage.2005.11.002>
- Visser M, Jefferies E, Lambon Ralph MA (2010) Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J Cogn Neurosci* 22(6):1083–1094
- Vitello S, Warren JE, Devlin JT, Rodd JM (2014) Roles of frontal and temporal regions in reinterpreting semantically ambiguous sentences. *Front Hum Neurosci* 8:530. <https://doi.org/10.3389/fnhum.2014.00530>
- Westerlund M, Kastner I, Al Kaabi M, Pylkkänen L (2015) The LATL as locus of composition: MEG evidence from English and Arabic. *Brain Lang* 141:124–134. <https://doi.org/10.1016/j.bandl.2014.12.003>
- Westerlund M, Pylkkänen L (2017) How does the left anterior temporal lobe contribute to conceptual combination? Interdisciplinary perspectives. In *Compositionality and concepts in linguistics and psychology* (S. 269–290). Springer, Cham.
- Whitney C, Kirk M, O’Sullivan J, Lambon Ralph MA, Jefferies E (2011) The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cereb Cortex* 21(5):1066–1075. <https://doi.org/10.1093/cercor/bhq180>
- Whitney C, Kirk M, O’Sullivan J, Lambon Ralph MA, Jefferies E (2012) Executive semantic processing is underpinned by a large-scale neural network: revealing the contribution of left prefrontal,

posterior temporal, and parietal cortex to controlled retrieval and selection using TMS. *J Cogn Neurosci* 24(1):133–147

Zaccarella E, Meyer L, Makuuchi M, Friederici AD (2015) Building by syntax: the neural basis of minimal linguistic structures. *Cereb Cortex*. <https://doi.org/10.1093/cercor/bhv234>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.