The Large-Scale Organization of Gestures and Words in the Middle Temporal Gyrus

Liuba Papeo,^{1*} Beatrice Agostini,^{2*} and Angelika Lingnau^{2,3,4}

¹Centre National de la Recherche Scientifique, Institut des Sciences Cognitives Marc Jeannerod, Unité Mixte de Recherche 5229 and University Claude Bernard Lyon 1, France, ²Center for Mind/Brain Science, University of Trento, 38068 Rovereto, Italy, ³Department of Psychology, Royal Holloway University of London, London, TW20 0EX United Kingdom, and ⁴Institute of Psychology, University of Regensburg, 93047 Regensburg, Germany

The middle temporal gyrus (MTG) has been shown to be recruited during the processing of words, but also during the observation of actions. Here we investigated how information related to words and gestures is organized along the MTG. To this aim, we measured the BOLD response in the MTG to video clips of gestures and spoken words in 17 healthy human adults (male and female). Gestures consisted of videos of an actress performing object-use pantomimes (iconic representations of object-directed actions; e.g., playing guitar), emblems (conventional gestures, e.g., thumb up), and meaningless gestures. Word stimuli (verbs, nouns) consisted of video clips of the same actress pronouncing words. We found a stronger response to meaningful compared with meaningless gestures along the whole left and large portions of the right MTG. Importantly, we observed a gradient, with posterior regions responding more strongly to gestures (pantomimes and emblems) than words and anterior regions showing a stronger response to words than gestures. In an intermediate region in the left hemisphere, the response was significantly higher to words and emblems (i.e., items with a greater arbitrariness of the sign-to-meaning mapping) than to pantomimes. These results show that the large-scale organization of information in the MTG is driven by the input modality and may also reflect the arbitrariness of the relationship between sign and meaning.

Key words: emblems; gestures; organization; pantomimes; speech

Significance Statement

Here we investigated the organizing principle of information in the middle temporal gyrus, taking into consideration the input-modality and the arbitrariness of the relationship between a sign and its meaning. We compared the middle temporal gyrus response during the processing of pantomimes, emblems, and spoken words. We found that posterior regions responded more strongly to pantomimes and emblems than to words, whereas anterior regions responded more strongly to words than to pantomimes and emblems. In an intermediate region, only in the left hemisphere, words and emblems evoked a stronger response than pantomimes. Our results identify two organizing principles of neural representation: the modality of communication (gestural or verbal) and the (arbitrariness of the) relationship between sign and meanings.

Introduction

Communication entails the mapping of signs, expressed through words or gestures, onto meanings. For more than a century, the temporal lobe has been considered as a central structure in the representation of meanings (Wernicke, 1874; Geschwind, 1970;

Price, 2012). Most available studies (for review, see Özyürek, 2014) have addressed either words (Kable et al., 2005; Peelen et al., 2012; Papeo and Lingnau, 2015) or gestures (Villarreal et al., 2008; Kubiak and Króliczak, 2016), and have consistently reported activity in the middle temporal gyrus (MTG) during understanding of both. A subset of studies has included both types of stimuli with the primary objective to identify the shared neural substrates (Xu et al., 2009; Andric et al., 2013). Once again, this research has highlighted activity in a brain network encompassing the MTG. The involvement in processing both words and gestures is consistent with the proposal that the entire MTG functions as a multimodal interface, interposed between the inferior temporal and the superior temporal gyri, specialized to processing of pictorial and verbal material, respectively (Binder et al., 2009; Visser et al., 2012; Wurm and Caramazza, 2019). While

Received Oct. 16, 2018; revised May 2, 2019; accepted May 5, 2019.

Author contributions: L.P. and A.L. designed research; L.P. and A.L. edited the paper; L.P. and A.L. wrote the paper; B.A. performed research; B.A. and A.L. analyzed data; B.A. wrote the first draft of the paper.

This work was supported by the Provincia Autonoma di Trento and the Fondazione Cassa di Risparmio di Trento e Rovereto. A.L. was supported by German Research Foundation Heisenberg-Professorship Grant Li 2840/2-1. L.P. was supported by European Research Council Starting Grant (Project THEMPO, Grant Agreement 758473).

The authors declare no competing financial interests.

*L.P. and B.A. contributed equally to this work.

Correspondence should be addressed to Angelika Lingnau at angelika.lingnau@ur.de.

<https://doi.org/10.1523/JNEUROSCI.2668-18.2019>

Copyright © 2019 the authors

Figure 1. Stimulus material and design. **A–D**, Still frames from video clips showing pantomimes (**A**; playing violin), emblems (**B**; thumb up), meaningless gestures (**C**), and spoken words (**D**). **E**, Experimental design and example trial. The experiment consisted of two parts (1: Gestures experiment; 2: Words experiment), using a block design. In the Gestures experiment, the conditions pantomimes, emblems (event and state), and meaningless gestures were presented. In the Words experiment, the conditions verbs (concrete and abstract) and nouns (concrete and abstract) were presented.

previous studies suggest that the MTG is implicated in representing the meaning of communicative signs, it is unclear how neural information is organized; in particular, whether neural representation of meanings in the MTG is abstracted away from the input modality, or rather reflects, in some way, whether meanings are conveyed through words or gestures.

While the input modality draws a sharp line between gestures and words, another property of communicative signs, the arbitrariness of the sign–meaning relationship, may entail another organization, where emblems are closer to words than pantomimes (McNeill, 2000). Pantomimes are iconic in that the sign reproduces or approximates the action, as it is performed in the real world (e.g., drinking from a glass, playing the guitar). Emblems, similarly to words, are characterized by a more arbitrary relationship between the sign (e.g., thumbs up, waving your hand) and its meaning (agreeing, greeting).

Here we asked to which degree the meaningfulness of a sign, its modality (gestural or verbal), and the relationship between sign and meaning contribute to the organization of information in the MTG. To this aim, we examined the BOLD response during the processing of meaningful and meaningless pantomimes and emblems and of spoken words, along the whole MTG strip. Overall, we found a stronger response to meaningful versus meaningless items along the whole strip. Moreover, the response to meaningful items showed a posterior-to-anterior gradient, with posterior regions responding more strongly to gestures than words and anterior regions responding more strongly to words than gestures. Moreover, an intermediate region along the strip in the left hemisphere showed stronger responses to words and emblems compared with pantomimes.

Materials and Methods

Participants

Seventeen native Italian-speaking participants (10 females; 7 males; mean age 24 years; age range 19–30 years) volunteered in the experiment. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric diseases. The study was approved by the ethics committee of the University of Trento. Participants gave written informed consent before participation.

Stimuli

The stimulus set consisted of 2.5-s-long video clips showing an actress who performed gestures silently, or spoken words (Fig. 1A–D). Gestures included the following: 60 pantomimes (iconic representations of object-directed actions, e.g., playing violin; Fig. 1A), 60 emblems (symbolic gestures, e.g., listening, thumb up; Fig. 1B), and 60 meaningless gestures (Fig. 1C). Emblems included two subsets: 30 emblems that referred to an action (emblems-event, e.g., to listen, to clap, to yawn) and 30 emblems that referred to a state (emblems-state, e.g., thumb up, no, victory). In all subsequent analyses (for details, see fMRI data analysis), we did not consider the distinction between emblems-event and emblems-state and collapsed the two subsets in one general stimulus group.

All gestures were selected from a publicly available database (<https://figshare.com/s/f6d27f6c213e38070842>) (Agostini et al., 2018), where a sample of 50 native Italians and 50 American raters judged the meaning of >200 gestures. For each stimulus (gesture), the database provides a measure of meaningfulness relative to a 1–7 point scale (in addition to the lexical entry most frequently associated with the gesture and the most frequent verbal description). For the current study, gestures with a median value of ≥ 5 in the meaningfulness score were included in the stimulus groups of meaningful gestures (pantomimes and emblems); gestures with a median value of ≤ 3 were included as meaningless gestures. Word stimuli spoken by the same actress who performed the gestures (Fig. 1D) were recorded using the same setting and apparatus of

gestures. Across all video-clips, the actress stood in front of the camera and pronounced each word aloud or produced a gesture. Words included 60 verbs and 60 nouns. Half of the verbs were concrete (e.g., *correre* — to run) and half were abstract (e.g., *pensare* — to think); likewise, half of the nouns were concrete (e.g., *casa* — house) and half were abstract (e.g., *talento* — talent). Verbs were presented in the singular first-person present tense (the first-person pronoun was always spoken; e.g., *io dipingo* — I paint), and nouns were presented in the singular form, preceded by the appropriate article (*la collina* — the hill). Concrete and abstract nouns and verbs were matched for length (number of graphemes) and frequency (Dizionario di frequenza della lingua italiana, Consiglio Nazionale delle Ricerche) (all p values > 0.05 ; pairwise t tests). In the subsequent analyses, we collapsed all concrete and abstract verbs and nouns in one general stimulus group for words, unless specified otherwise (for details, see fMRI data analysis).

Inside the scanner, visual stimuli were back-projected onto a screen (frame rate: 60 Hz; screen resolution: 1024×768 pixels) via a liquid crystal projector (OC EMP 7900; Epson) and viewed through a mirror mounted on the head coil. Sounds were presented via MR-compatible headphones (SereneSound Digital audio system). Stimulus presentation, response collection, and synchronization with the scanner were controlled with ASF (Schwarzbach, 2011) and the MATLAB Psychtoolbox-3 for Windows (Brainard, 1997).

Design

The experiment consisted of two parts, using a block design (Fig. 1E). In the first part (Gestures experiment), consisting of six runs, participants were presented with blocks of pantomimes, emblems, and meaningless gestures. Half of the blocks for the emblems condition were emblems-state and half of the blocks were emblems-event, but these were modeled as one single condition (emblems). In the second part (Words experiment), consisting of two runs, participants were presented with blocks of verbs and nouns. Half of the blocks of verbs and nouns were concrete and half of the blocks were abstract; but in the main analysis, we collapsed across all concrete and abstract nouns and verbs (see Results).

Each block included five videos of the same experimental condition, separated by 1 s of fixation (Fig. 1E). Each run consisted of 12 blocks, separated by 16.5 s of fixation, and lasted ~ 6.9 min. Each run included the same number of blocks per condition. The Gestures experiment included 4 blocks for each condition (pantomimes, emblems, and meaningless gestures). The Words experiment included six blocks for each condition (nouns, verbs). Each video was presented only once per run, except for one video per run that was repeated twice, to allow a one-back task (see Task). The order of blocks within each run was randomized. Each fMRI session ended with a movement localizer run, using a block design (15 blocks; 16 s duration; 16.5 s fixation period between blocks), in which participants moved the right hand, the right foot, or the tongue (5 blocks per conditions), following written instructions presented on the screen. The movement localizer was collected as part of another study and therefore will not be discussed further.

Task

Participants were instructed to watch each video clip, and to press the response button with the left index finger whenever a stimulus was repeated twice in a row (one-back task). To avoid a systematic association between the button press and one of the experimental conditions (and thus the corresponding BOLD response), there was only one repetition per run, randomly assigned to one of the conditions.

Data acquisition

Functional and structural data were acquired using a 4T Bruker MedSpec Biospin MR scanner and an 8 channel birdcage head coil. Functional images were acquired with a T2*-weighted gradient EPI sequence. Acquisition parameters were as follows: TR of 2000 ms; TE of 33 ms; voxel resolution of $3 \times 3 \times 3$; flip angle of 73° ; FOV of 192×192 mm; gap size of 0.45 mm. We used 28 slices, acquired in ascending interleaved order. In each functional run, 207 images were acquired. Before each functional run, we performed an additional scan to measure the point-spread function (PSF) of the acquired sequence to correct the distortion expected with high-field imaging (Zaitsev et al., 2004). Structural T1-weighted

anatomical scans were acquired with an MPRAGE sequence (176 sagittal slices; TR 2700 ms; TE 4.18 ms; voxel resolution $1 \times 1 \times 1$ mm; flip angle 7° ; FOV 256×224 mm; inversion time 1020 ms) to coregister low-resolution functional images with a high-resolution anatomical scan.

fMRI data analysis

Preprocessing. Data were preprocessed and analyzed using BrainVoyager QX 2.8 (Brain Innovation) in combination with the BVQtools/NeuroElf toolbox (by Jochen Weber; <http://neuroelf.net/>) and MATLAB (The MathWorks). Distortion in geometry and intensity in the EPI images was corrected on the basis of the PSF data acquired before each functional run (Zeng and Constable, 2002). The first 4 volumes were discarded to avoid T1 saturation. The first volume of the first functional run was aligned to the high-resolution anatomy using 6 rigid-body transformation parameters. 3D motion correction (trilinear interpolation) was performed using the first volume of the first run of each participant as reference, followed by slice time correction and high-pass filtering (3 cycles per run). Spatial smoothing was applied with a Gaussian kernel of 5 mm FWHM. For group analyses, both functional and anatomical data were transformed into Talairach space, using trilinear interpolation.

Univariate analysis. We used a random-effects (RFX) GLM analysis, including regressors for the conditions pantomimes, emblems, meaningless gestures, and spoken words. To assess the sensitivity of the MTG to meaningful stimuli, we included the RFX GLM contrast meaningful $>$ meaningless gestures. To assess the sensitivity of the MTG to gestures and words, we computed RFX GLM contrasts for gestures (pantomimes, emblems) and words against the implicit baseline. Statistical maps were corrected using Threshold Free Cluster Enhancement (TFCE) for cluster-level correction (corrected cluster threshold: $\alpha = 0.05$, using Monte Carlo simulations with 10,000 permutations) as implemented in CoSMoMvPA (Oosterhof et al., 2016). Specific differences across categories of meaningful stimuli were addressed in the subsequent ROI analysis with greater statistical power.

Posterior-to-anterior organization. To investigate the organization of information in the MTG, we assessed whether the amplitude of the BOLD response to our main experimental conditions (pantomimes, emblems, words) varied as a function of the y coordinate along the MTG. To this end, we identified the left and right MTG on the segmented and inflated reconstructions of the Talairach-transformed version of the MNI-Colin 27 template based on the following anatomical landmarks: the superior and inferior borders corresponded to the superior temporal sulcus and the inferior temporal sulcus, respectively (Fig. 2, top); the posterior end was delineated by the preoccipital notch and extended anteriorly along the sagittal plane for the entire length of the temporal lobe. The entire ROI encompassed 17,523 voxels.

For each voxel in the MTG ROI, we extracted β estimates of every experimental condition, separately for each participant. The design matrix and predictors were the same used for computing the RFX GLM for the univariate analysis (see fMRI data analysis: Univariate analysis). Beta estimates were averaged across the x and z dimensions to obtain one value for each y coordinate (Fig. 2, bottom).

To identify the spatial positions along the y axis in which any of the contrasts of interest showed a significant difference, we computed paired t tests at each y coordinate, correcting for the number of voxels using TFCE (Smith and Nichols, 2009) as implemented in CoSMoMvPA (Oosterhof et al., 2016). In brief, this involved the following steps, performed separately for each of the contrasts of interest. (1) For each voxel, a cluster was defined based on its proximal neighboring voxels (i.e., the previous and the following one); (2) a univariate group-level analysis (ANOVA) was computed across all the conditions for the contrast of interest (see Results), and an uncorrected F map was obtained; (3) the F map was converted into a z map; (4) a TFCE map was computed from the z map obtained before, using the recommended values of $h_0 = 0$, $E = 0.5$, $H = 2$, and $dh = 0.1$ (for details, see Smith and Nichols, 2009); (5) a null-distribution was generated performing a group-level analysis for 10,000 iterations (using Monte Carlo simulations); (6) at each iteration, a TFCE map was computed from the obtained z map, and the maximum value was saved for later; and (7) the corrected TFCE map was obtained by counting how often, for each voxel, the observed TFCE value was

Figure 2. Top, The MTG ROI was defined as the cortical region inferior to the superior temporal sulcus (STS) and superior to the inferior temporal sulcus (ITS). Bottom, Procedure used for the MTG ROI analysis.

smaller than the maximum (see Step 5) TFCE value for all the null TFCE maps. The resulting number was then divided by the number of iterations to obtain a corrected z map that allowed determining which clusters survived the correction for multiple comparisons.

Results

Univariate analysis

To identify the brain network sensitive to the distinction between meaningful versus meaningless stimuli, we computed the whole-brain RFX GLM contrast Meaningful > Meaningless Gestures (Fig. 3A). As can be seen, most of the left MTG and the most anterior portion of the right MTG are sensitive to this difference.

To get a first hint at the areas recruited during the processing of gestures and words, we first computed the whole-brain RFX GLM contrasts Gestures > Baseline (Fig. 4A) and Words > Baseline (Fig. 4B). Both contrasts revealed a bilateral recruitment of the temporal cortex. Moreover, the contrast Gestures > Baseline recruited the posterior part of the MTG, as well as several parietal (BA7/BA40) and inferior and middle frontal (BA44/45/46/47) areas. The contrast Words > Baseline revealed clusters in the middle and anterior portion of the MTG, the superior temporal gyrus, as well as in the inferior and middle frontal gyrus.

Posterior-to-anterior organization of the MTG

To examine the posterior-to-anterior organization of the MTG, we plotted the β estimates for each condition as a function of the y coordinate within this ROI (for details, see Materials and Methods; Fig. 2).

Meaningful versus meaningless gestures

First, we focused on gestures to test the sensitivity of the MTG to the broad distinction between meaningful and meaningless items. In line with the results of the corresponding whole-brain RFX GLM contrast (Fig. 3A), we obtained a higher amplitude of the BOLD signal for meaningful compared with meaningless items, both in the left and right hemisphere, although the effect was stronger in the left hemisphere (Fig. 3B–D).

In the left hemisphere, meaningful items led to a higher BOLD amplitude than meaningless items along the entire MTG (Fig. 3B, left). This effect was replicated with the contrasts pantomimes versus meaningless gestures (Fig. 3C, left) and emblems versus meaningless gestures (Fig. 3D, left).

In the right hemisphere, the comparison between meaningful and meaningless gestures was significant at anterior and posterior positions, but not at an intermediate y position ~ -20 (Fig. 3B, right). Identical results were found with the contrast pantomimes versus meaningless gestures (Fig. 3C, right), whereas the contrast emblems versus meaningless gestures showed a difference only in the middle ($y = \sim -30$) and anterior ($y = \sim 0$) part of the right ROI (Fig. 3D, right).

In sum, using the case of gestures, the above analyses revealed that the left and right MTG show a consistently higher amplitude of the BOLD signal to meaningful than meaningless items.

Meaningful gestures versus words

In the following analysis, we included the neural responses to words to study the organization of meaningful information in the MTG according to input modality and/or arbitrariness of the sign–meaning relation. To this aim, we plotted the β estimates for each meaningful condition (pantomimes, emblems, and words) as a function of the y position within the MTG ROI (Fig. 4C). We observed a structured organization of information along the posterior-to-anterior axis, whereby posterior regions responded more strongly to gestures than words, and anterior regions responded more strongly to words than gestures. This pattern held both in the left (Fig. 4C, left) and the right (Fig. 4C, right) hemisphere. In an intermediate region in the left hemisphere ($y = \sim -30$), emblems and words showed comparable activity, which was stronger than the activity for pantomimes (Fig. 4A, left).

Statistical analyses supported this description (Fig. 4C, gray bars illustrating positions along the y axis at which statistical comparisons survived TFCE correction). Pantomimes and emblems compared with words yielded a significantly higher BOLD response in the pos-

Figure 3. Sensitivity to meaningful compared with meaningless gestures. **A**, Results of the whole-brain RFX GLM contrast meaningful versus meaningless gestures. **B**, BOLD amplitude (expressed as β estimates) to meaningful (black) and meaningless (blue) gestures as a function of the y position along the left (left column) and right (right column) MTG. **C**, BOLD amplitude for pantomimes and meaningless gestures. **D**, BOLD amplitude for emblems and meaningless gestures. Gray bars represent clusters that survived correction for multiple comparisons using TFCE (for details, see Materials and Methods).

terior MTG, both in the left and right hemisphere. By contrast, words compared with pantomimes and emblems yielded a stronger BOLD response in the anterior MTG bilaterally.

The gradients for words, with higher activity in anterior than in posterior regions, and for gestures (i.e., pantomimes and emblems), with higher activity in posterior than in anterior regions, were obtained not only in relation to the other stimulus group. Compared against baseline, word-related ac-

tivity was significantly higher at more anterior sites (between $y = -16$ and -46), whereas gesture-related activity was significantly above baseline at the more posterior sites (between $y = -36$ and -70), in line with the results of the corresponding whole-brain RFX GLM contrasts (Fig. 4*A, B*). Finally, we note that activity for each stimulus group (pantomimes, emblems, and words) at each ROI along the y axis did not differ statistically between the left and right hemisphere.

Figure 4. Effects of different categories of meaningful stimuli. **A, B**, Results of the whole-brain contrasts Gestures versus Baseline (**A**) and Words versus Baseline (**B**). **C**, Beta estimates for pantomimes (red), emblems (green), and words (yellow) as a function of the y coordinate along the left (left) and right (right) MTG. Gray bars represent clusters that survived corrections for multiple comparisons using TFCE (for details, see Materials and Methods) for each of the pairwise comparisons; light gray represents pantomimes versus emblems; medium gray represents emblems versus words; dark gray represents pantomimes versus words.

Pantomimes versus emblems

Considering emblems and pantomimes separately, we found stronger activity to the former in an intermediate region along the left MTG ($y \sim -30$) (Fig. 4C, left). Pantomimes compared with emblems yielded stronger activity in the right posterior MTG (Fig. 4C, right). In the same intermediate region that showed a significantly stronger BOLD signal for emblems than for pantomimes ($y \sim -30$; Fig. 4C, left), emblems and words elicited comparable responses.

In addressing the difference between pantomimes and emblems, we considered that the former stimulus group only denoted concrete object-directed actions, whereas the latter implied both concrete actions and more abstract contents (and so did words). Thus, regions where pantomimes and emblems differed could capture the distinction between abstract and concrete meanings. To evaluate this possibility, we compared the activity in the left and right MTG, for length- and frequency-matched abstract and concrete verbs, and abstract and concrete nouns. We obtained no significant difference between abstract and concrete

verbs (Fig. 5A), whereas we found a higher amplitude of the BOLD signal for abstract versus concrete nouns in a cluster $y \sim -26$ in the right MTG (Fig. 5B, right), outside the region where emblems differ from pantomimes. Weak effects of abstractness/concreteness are compatible with current results showing comparable activity in the lateral temporal cortex for abstract and concrete words (Bedny et al., 2008; Hernández et al., 2014; Papeo and Lingnau, 2015), and abstractness/concreteness effects in regions outside this territory (Binder et al., 2009; Straube et al., 2013). In summary, although emblems and pantomimes elicited comparable activity in the largest part of left and right MTG, they differed in two clusters. We have ruled out the possibility that this difference is accounted for by a difference in the abstractness/concreteness of meanings implied by the two categories of gestures.

Discussion

The MTG is consistently recruited during both action observation and the processing of words. Posterior regions of the MTG

Figure 5. Comparison of abstract and concrete words. **A**, Comparison of the BOLD amplitude for concrete and abstract verbs in the left (left) and right (right) hemisphere. **B**, Comparison of the BOLD amplitude for concrete and abstract nouns in the left (left) and right (right) hemisphere. Gray bars represent clusters that survived corrections for multiple comparisons using TFCE (for details, see Materials and Methods).

have been shown to be preferentially recruited during the processing of meaningful visual actions (Caspers et al., 2010; Wurm and Lingnau, 2015; Wurm et al., 2016, 2017), whereas anterior regions are specially recruited during language processing (Kable et al., 2005; Chatterjee, 2008; Watson et al., 2013; Weiner and Grill-Spector, 2013; Lingnau and Downing, 2015; Tarhan et al., 2015). However, anatomical and functional relations between action and language/speech processing are harder to draw because extant studies have addressed either one or the other. Moreover, among visual actions, object-use actions (e.g., slicing bread) have been investigated more extensively than communicative acts (Kable et al., 2005; Chatterjee, 2008; Watson et al., 2013; Weiner and Grill-Spector, 2013; Tarhan et al., 2015). In previous studies that included language/speech and communicative gestures in the same design, the analyses addressed the commonalities (Xu et al., 2009; Andric et al., 2013) as opposed to defining the large-scale organization of categories of meaningful signs. Here we performed a systematic analysis of the organization of meanings in the MTG. We considered both the role of the modality of the sign through which a meaning is conveyed (verbal vs gestural), and the arbitrariness of the sign-meaning relationship, which, on a virtual continuum, places emblems closer to words than pantomimes (McNeill, 2000). Our results showed a structured organization along the posterior-to-anterior axis, where posterior regions responded more strongly to meaningful gestures (pantomimes, emblems) than words, and anterior regions showed a stronger response to words than meaningful ges-

tures. In an intermediate region in the left hemisphere (y coordinate = ~ -30 , BA21), we found a significantly higher BOLD amplitude for words and emblems (i.e., stimuli with an arbitrary sign-to-meaning mapping) compared with pantomimes. In the following, we will discuss these results in depth.

Meaningful and meaningless gestures

We found that the whole MTG responded more strongly to meaningful than meaningless gestures, in keeping with a century of research on the role of the temporal lobe in representation of meanings (Wernicke, 1874; Geschwind, 1970; see also Kable et al., 2002, 2005; Noppeney, 2008; Villarreal et al., 2008; Kalénine et al., 2010; Whitney et al., 2011; Peelen et al., 2012; Visser et al., 2012; Papeo and Lingnau, 2015; Papeo et al., 2015).

An intermediate region of the left MTG showed a significantly stronger BOLD signal in response to emblems compared with pantomimes. Relative to emblems, pantomimes evoked a stronger BOLD signal in the right posterior MTG. Since pantomimes implied only concrete meanings (object-use actions) while emblems included both concrete and more abstract meanings, we asked whether the difference between the two was driven by a difference in semantic abstractness/concreteness. An additional analysis, in which we compared abstract and concrete words along the MTG, ruled out this possibility: the effect of abstractness/concreteness was only found in a small cluster outside the regions that distinguished between pantomimes and emblems. Another source of difference between pantomimes and emblems

is transitivity (i.e., the object-directedness of an action), a property linked to posterior occipitotemporal activity (Wurm et al., 2017). While transitivity could account for the higher response to pantomimes than emblems in the right posterior MTG, it is less likely to account for the higher response to emblems than pantomimes in the middle portion of the left MTG.

Both pantomimes and emblems are communicative gestures, but, differently from pantomimes and similarly to words, emblems are characterized by an arbitrary relation between sign and meaning. By virtue of such arbitrariness, it has been suggested that emblems incorporate linguistic properties and are represented in a mental lexicon (McNeill, 2005). The abovementioned left-lateralized difference between emblems and pantomimes could reflect differences in the sign-to-meaning mapping, linked to the arbitrariness of the relationship between the two. This possibility is encouraged by the fact that, in the same MTG site, neural activity was comparable for emblems and words (see below).

Meaningful stimuli: pantomimes, emblems, and words

Moving along the posterior-to-anterior axis, the neural response to pantomimes and emblems decreased, whereas the response to words increased, with a slight decrease in the most anterior portions, both in the left and right hemisphere (Fig. 4C). At an intermediate location in the left MTG ($y \sim -30$), the activity for words was stronger than for pantomimes, and comparable with emblems. More anteriorly, words induced stronger activity relative to both emblems and pantomimes.

One might wonder whether, being in a territory that cares about semantic content, the current anteroposterior organization is driven by the meanings of stimuli, somehow confounded with the modality (gestures or words). Pantomimes only implied concrete object-use actions, whereas words and emblems carried both concrete and abstract contents. If semantic distinctions, coarsely captured by the concrete/abstract distinction, were the only responsible of this organization, pantomimes should have differed the most from the other stimulus groups. Instead, in the largest part of the left MTG, activity for pantomimes and emblems was comparable, and different from activity for words. Moreover, the effect of abstractness/concreteness, tested through the comparison between abstract versus concrete words, was found in a circumscribed right anterior region. This may account for the difference between pantomimes and emblems in the right MTG (found posterior to this region), but not for the relation between words and gestures in the whole bilateral MTG. Semantic distinctions beyond the abstract-concrete one may contribute to the organization of information in the MTG. A growing corpus of studies is unraveling neural representational spaces nestled in various sectors of the MTG, organized according to various semantic or perceptual dimensions and/or types of processes (Lingnau and Downing, 2015; Wurm et al., 2017). These spaces could form a meso-organization within the large-scale, macro-organization found here, symmetrically in both hemispheres.

As argued for the organization of object representations in the inferior temporal cortex (Mahon and Caramazza, 2011), the functional organization of cortical areas might be driven by systematic differences in the underlying connectivity patterns. Could a similar principle underlie the large-scale organization of meaningful information that we obtained in the MTG? On this hypothesis, more anterior portions of the MTG could show connectivity profiles that are specific to the processing of communication via speech (e.g., auditory stimulation, phonemic and

phonological analysis), whereas more posterior portions might show connectivity with areas involved in communication via gestures (e.g., processing of kinematics). Support for this idea comes from Turken and Dronkers (2011) who examined structural and functional connectivity in healthy participants, in a number of regions that had been shown to be critical for language comprehension in previous patient studies. Moving from posterior to anterior sections of the MTG, both resting-state functional connectivity analysis and diffusion tensor imaging-based connectivity revealed an increase in connectivity with frontal regions (particularly with the perisylvian territory encompassing the inferior frontal gyrus) and a decrease in connectivity with superior temporal and parietal regions. In summary, sections 1 and 4 of the Turken and Dronkers (2011) dataset, corresponding to the regions preferring words and gestures, respectively, in the current study, were most dissimilar in terms of connectivity profile, whereas sections 2 and 3, where we obtained similar responses to emblems and words, showed intermediate connectivity profiles incorporating those seen in segments 1 and 4.

Our results are also in line with an activation likelihood estimation meta-analysis on the processing of action concepts (Watson et al., 2013). This analysis revealed a consistent recruitment, across studies, of the lateral occipital cortex for object-use actions relative to words, and a consistent recruitment for action words relatively to object-use actions in the middle portion of MTG. The conjunction of object-use actions and action words revealed the posterior MTG (anterior to the lateral occipital cortex and posterior to the middle portion of the MTG). We emphasize that modality effects in the MTG reflect fine-grained distinctions underneath the general meaningful versus meaningless distinction. Primarily, the MTG responds more strongly to meaningful compared with meaningless stimuli; as such, it is a multimodal area (Binder et al., 2009; Visser et al., 2012; Wurm and Caramazza, 2019).

In conclusion, our results contribute to clarify an architecture with pantomimes, emblems, and words organized along a continuum of communicative signs that differ in terms of input modality (visual for pantomimes and emblems vs verbal-auditory for words) and arbitrariness of the relationship between input and meaning (lower for pantomimes, greater for emblems and words). This architecture emerges from the organization of information in the MTG, where posterior regions represent gestural signs (pantomimes, emblems), and anterior regions represent verbal signs (words). An intermediate site along this structure responds differently to words and emblems versus pantomimes. This site might capture the difference in the sign-to-meaning mapping, reflecting the arbitrariness of the relationship between the two. In conclusion, our study shows that the MTG is sensitive to the difference between meaningful and meaningless stimuli and is characterized by an internal organization, where pantomimes, emblems, and words are represented along a posterior-to-anterior gradient. This gradient may reflect the sign-to-meaning relationship of stimuli that is characterized by a gradual decrease of concrete sensorimotor properties and an increase of abstract linguistic properties.

References

- Agostini B, Papeo L, Galusca CI, Lingnau A (2018) A norming study of high-quality video clips of pantomimes, emblems, and meaningless gestures. *Behav Res Methods*. Advance online publication. Retrieved December 12, 2018. doi: 10.3758/s13428-018-1159-8.
- Andric M, Solodkin A, Buccino G, Goldin-Meadow S, Rizzolatti G, Small SL (2013) Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51:1619–1629.

- Bedny M, Caramazza A, Grossman E, Pascual-Leone A, Saxe R (2008) Concepts are more than percepts: The case of action verbs. *J Neurosci* 28:11347–11353.
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Caspers S, Zilles K, Laird AR, Eickhoff SB (2010) ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* 50:1148–1167.
- Chatterjee A (2008) The neural organization of spatial thought and language. *Semin Speech Lang* 29:226–238; quiz C6.
- Geschwind N (1970) The organization of language and the brain. *Science* 170:940–944.
- Hernández M, Fairhall SL, Lenci A, Baroni M, Caramazza A (2014) Predication drives verb cortical signatures. *J Cogn Neurosci* 26:1829–1839.
- Kable JW, Lease-Spellmeyer J, Chatterjee A (2002) Neural substrates of action event knowledge. *J Cogn Neurosci* 14:795–805.
- Kable JW, Kan IP, Wilson A, Thompson-Schill SL, Chatterjee A (2005) Conceptual representations of action in the lateral temporal cortex. *J Cogn Neurosci* 17:1855–1870.
- Kalénine S, Buxbaum LJ, Coslett HB (2010) Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. *Brain* 133:3269–3280.
- Kubiak A, Króliczak G (2016) Left extrastriate body area is sensitive to the meaning of symbolic gesture: evidence from fMRI repetition suppression. *Sci Rep* 6:31064.
- Lingnau A, Downing PE (2015) The lateral occipitotemporal cortex in action. *Trends Cogn Sci* 19:268–277.
- Mahon BZ, Caramazza A (2011) What drives the organization of object knowledge in the brain? *Trends Cogn Sci* 15:97–103.
- McNeill D (2000) Language and gesture, Vol 2. Cambridge: Cambridge UP.
- McNeill D (2005) Gestures and thoughts. Chicago: University of Chicago.
- Noppeney U (2008) The neural system of tool and action semantics: a perspective from functional imaging. *J Physiol Paris* 102:40–49.
- Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMPPA: multimodal multivariate pattern analysis of neuroimaging data in Matlab / GNU Octave. *Front Neuroinform* 10:27.
- Özyürek A (2014) Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos Trans R Soc Lond B Biol Sci* 369:20130296.
- Papeo L, Lingnau A (2015) First-person and third-person verbs in visual motion-perception regions. *Brain Lang* 141:135–141.
- Papeo L, Lingnau A, Agosta S, Pascual-Leone A, Battelli L, Caramazza A (2015) The origin of word-related motor activity. *Cereb Cortex* 25:1668–1675.
- Peelen MV, Romagno D, Caramazza A (2012) Independent representations of verbs and actions in left lateral temporal cortex. *J Cogn Neurosci* 24:2096–2107.
- Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62:816–847.
- Schwarzbach J (2011) A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behav Res Methods* 43:1194–1201.
- Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44:83–98.
- Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G, Nagels A (2013) Supramodal neural processing of abstract information conveyed by speech and gesture. *Front Behav Neurosci* 7:120.
- Tarhan LY, Watson CE, Buxbaum LJ (2015) Shared and distinct neuroanatomic regions critical for tool-related action production and recognition: evidence from 131 left-hemisphere stroke patients. *J Cogn Neurosci* 27:2491–2511.
- Turken AU, Dronkers NF (2011) The neural architecture of the language comprehension network: converging evidence from lesion connectivity analyses. *Front Syst Neurosci* 5:1.
- Villarreal M, Fridman EA, Amengual A, Falasco G, Roldan Gerscovich ER, Ulloa ER, Leiguarda RC (2008) The neural substrate of gesture recognition. *Neuropsychologia* 46:2371–2382.
- Visser M, Jefferies E, Embleton KV, Lambon Ralph MA (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J Cogn Neurosci* 24:1766–1778.
- Watson CE, Cardillo ER, Ianni GR, Chatterjee A (2013) Action concepts in the brain: an activation likelihood estimation meta-analysis. *J Cogn Neurosci* 25:1191–1205.
- Weiner KS, Grill-Spector K (2013) Neural representations of faces and limbs neighbour in human high-level visual cortex: evidence for a new organization principle. *Psychol Res* 77:74–97.
- Wernicke C (1874) Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis. Berlin: Cohn and Weigert.
- 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* 21:1066–1075.
- Wurm MF, Caramazza A (2019) Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nat Commun* 10:289.
- Wurm MF, Lingnau A (2015) Decoding actions at different levels of abstraction. *J Neurosci* 35:7727–7735.
- Wurm MF, Ariani G, Greenlee MW, Lingnau A (2016) Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cereb Cortex* 26:3390–3401.
- Wurm MF, Caramazza A, Lingnau A (2017) Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *J Neurosci* 37:562–575.
- Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR (2009) Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci U S A* 106:20664–20669.
- Zaitsev M, Hennig J, Speck O (2004) Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn Reson Med* 52:1156–1166.
- Zeng H, Constable RT (2002) Image distortion correction in EPI: comparison of field mapping with point spread function mapping. *Magn Reson Med* 48:137–146.