An exploratory fMRI study on metonymy and metaphor processing

Sofia Fregni[1], Karin Heidlmayr[2], Kirsten Weber[2], & David Peeters[2, 3]

[1]*Donders Institute for Brain, Cognition, and Behavior, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, 6500 HB Nijmegen, The Netherlands,* [2]*Max Planck Institute for Psycholinguistics, 6500 AH, Nijmegen, The Netherlands, and* [3]*Tilburg School of Humanities and Digital Sciences, Language, Communication, and Cognition, 5037 AB Tilburg, The Netherlands*

ABSTRACT

Metaphor and metonymy are believed to play a key role in human communication, language, and cognition. However, it remains unclear what neural mechanisms allow for their comprehension, and whether the brain regions involved during their processing reflect enhanced linguistic processing or inferential processes. Through the use of fMRI and two localizer tasks, for language and Theory-of-Mind, we investigated the mechanisms underlying the comprehension of metaphoric and metonymic language. Both types of figurative language elicited increased left-lateralized frontotemporal activity, with overlaps across regions belonging to both the Theory-of-Mind network and the language network. Metonymy processing additionally recruited the right inferior frontal gyrus, whereas metaphor comprehension downregulated the frontoparietal control system. Results are discussed in terms of the degree of semantic distance between the metaphoric and metonymic allusions and their referent.

Keywords: metaphor, metonymy, figurative language, theory-of-mind, cognitive control

1. Introduction

Everyday language is remarkably rich in nonliteral sentences. Figurative language is one of the most widely recognized forms of *conversational implicature*, that is the act of meaning or implying one thing by saying something else (Wayne, 2014). I may ask a friend to have a *glass* with me, implying that I would like her to drink wine with me. Similarly, I may describe a person as *explosive*, while referring to his temperament. The interest in conversational implicatures has brought semantics and pragmatics to jointly attempt and shed light on the mechanisms supporting speakers' meaning construction and hearers' meaning selection. As the previous examples show, in figurative language, the speaker's meaning goes beyond what she literally says. According to Bach (1994), "nonliteral language has both a literal meaning that is tied to the semantics of the words, and a figurative one that can be derived from it". It follows that we need to distinguish one meaning from the other and pick the one that better fits the contextual cues that social communication provides us with. Inferential pragmatic processes would then be the means through which hearers can correctly interpret speakers' *intended* meaning and thus close the gap between this and the semantics of the words that are uttered (Bach & Harnish, 1979; Eizaga Rebollar, 2015).

The present study focuses on the mechanisms underlying the hearer's understanding of nonliteral language. In particular, we will focus on metaphor and metonymy as two subtypes of nonliteral language use. The interest in these two types of language use stems from their ubiquity in everyday language (across languages and communication modes, Mio & Katz, 1996; Boerger, 2005), and the important role they play in human communication (for a discussion see Lakoff, 1987; Lakoff & Johnson, 1980; Boerger, 2005).

1.1. Metonymy and metaphor: are they the same?

Metonymy is defined as using a particular property of something or its special relationship with some other thing to refer to it (Song, 2011). It follows that metonymy is the means through which the speaker can help the hearer to locate or to recognize the referent and its particular properties (Song, 2011). Let John be sat in a restaurant, say at table 5, and let him order a steak. Once he finishes eating, the waiter approaches John and asks him whether he desires something more. John asks for the bill. Now, let the waiter go to the check-out and say: “Table 5 wants to pay the bill”. In this example, the waiter uses ‘table 5’ to refer to the customer sat there, i.e. John, rather than attributing desires to a table (for a discussion, see Nunberg, 1979; Jackendoff, 1997). The plausibility to use ‘table 5’ to refer to the customer is rooted in the relationship of contiguity between the costumer and his presence at that table (Lakoff, 1987; Lakoff and Turner, 1989), where contiguity is defined as conceptual of physical vicinity. If John were wearing a flashy yellow coat, the waiter might have called him ‘the flashy yellow coat’ instead, still on the basis of a perceived contiguity between the referent and some property of the referent used to allude to it.

Metaphoric language, on the other hand, allows for explaining one thing *in terms* of another. Let now John be in the restaurant with Anne. And let Anne utter to John while he is voraciously eating: "You are a pig!". Anne is surely not speaking literally. Rather, she is implying that John is greedy and eager. In this example, the plausibility of Anne's utterance cannot be explained by a contiguity mechanism anymore. Whereas the relationship between John and his being sat at table 5, or his wearing a flashy yellow coat, are salient characteristics belonging to him, the relationship between John and a pig is to a large extent arbitrary. One could interpret John as a pig because he is greedy, just as because he is overweight. That is, there is not an interpretation of John’s *piggishness* that is better than the other. Given the distance between the conceptual domain of a person, in this case John, and the conceptual domain of an animal, here a pig, it is up to the hearer to infer why John could be so. Accordingly, metaphor relies on the mapping between a target and a source domain, where the target, i.e. what the speaker wants to explain, and the source, i.e. the means by which the target is explained, are conceptually distant, and thus, their mapping depends on the perceived similarity experienced by speakers themselves.

From the theoretical point of view, metaphor and metonymy appear to be different mechanisms of language use. Accordingly, whereas both types of language rely on the mapping between a source and a target domain, this differs in terms of quantity (metaphoric mapping is between two unrelated conceptual domains, whereas metonymic one stays within the same domain), as well as from a qualitative perspective, i.e. metonymic contiguity is different from metaphoric similarity. Yet, it is an open question whether the underlying mechanism for processing these two types of figurative language are the same or whether they differ. Neuroscience in the last decade contributed significantly to the understanding of metaphor and metonymy processing. Through the use of methods from cognitive neuroscience, it became possible to directly investigate the neurocognitive mechanisms involved in processing different types of figurative language.

In the context of electrophysiology, the ERP literature consistently reports two components during metaphor comprehension, the N400 and the late positivity LP/ P600. The N400 has been traditionally thought to reflect false expectation and prediction errors, as well as semantic integration and processing ease (for a review, see Kutas & Federmeier, 2011). The Late Positivity, on the other hand, although still controversial, has been interpreted as mirroring discourse updating, enriched composition, and pragmatic processes. With respect to metaphor, these components are thought to reflect different stages of processing and conceptual mapping between the target and the source domain. However, the literature divides into studies reporting a N400-LP biphasic pattern (e.g. De Grawe et al., 2010; Weiland et al., 2014), and others only a monophasic pattern, either the N400 or the LP component (Pynte et al., 1996; Yang et al., 2013, respectively). Less studied than metaphor, metonymy only recently gained more interest. However, just like for metaphor, some studies reported a biphasic N400-LP pattern (Schumacher, 2014), while others only a monophasic N400 (Weiland et al., 2014) or LP pattern (Schumacher, 2013). One hypothesis is that the inconsistency in the ERP findings for both metaphor and metonymy comprehension might be related to experimental designs that did not control for the variable of context during the comprehension of the stimuli. Accordingly, context, when supportive for understanding nonliteral stimuli, seems to play a key role in suppressing the N400 (for a discussion see Bambini et al., 2016). Yet, the inconsistency in the ERP literature for metaphor and metonymy comprehension leaves a window open to the extent to which these two types of figurative language might be supported by the same underlying mechanism. The only comparison, albeit indirect, has been attempted in 2014 by Weiland and colleagues, who paralleled metaphor to metonymy while measuring participants' brain activity. Metaphor comprehension elicited a biphasic N400-LP pattern, whereas metonymy processing only a monophasic N400. The lack of Late Positivity for metonymy has been taken by the authors as suggesting a lower cognitive cost for this type of figurative language in comparison to metaphor. The stimuli they used, however, were not matched between metaphoric and metonymic conditions. Moreover, the only type of metonymy they used was the producer-for-product type, hence it remains unclear whether their findings can account for all metonymic language use. Producer-for-product metonymy is a conventionalized way of referring to someone's work via his name (for example in ‘I love reading Shakespeare'). Although Weiland et al. controlled for the familiarity of the famous person used in the stimuli, the conventionality of adopting this type of association in everyday language might have facilitated participants toward the correct selection of the nonliteral meaning. This, in turn, might explain the absence of further pragmatic processes, i.e. Late Positivity, in the metonymic condition. Accordingly, in Schumacher (2014), nonconventional metonymies elicited a biphasic N400- LP pattern similar to the one that Weiland et al. found for metaphoric sentences.

With respect to the neural architecture underlying metaphor and metonymy processing, the fMRI literature shows a partial overlap between metaphor and metonymy comprehension underlying mechanisms. Rapp et al. (2012) meta-analyzed sixteen fMRI studies for the contrast metaphor > literal at the whole-brain level, and identified the strongest significant clusters of activation in the left parahippocampal gyrus (BA 36) and left inferior frontal gyrus (BA 47). Moreover, their analysis showed significant clusters in both hemispheres (BA 47, 34 and 21), although, overall, a left lateralization was still present. Other significant clusters were found in the middle frontal gyrus (BA 9), middle temporal gyrus (BA 22), superior occipital gyrus (BA 19), medial frontal gyrus (BA 9), inferior parietal lobule (BA 40), middle temporal gyrus (BA 21), thalamus, and anterior cingulate (BA 24) in the left hemisphere. Another meta-analysis (Bohrn et al., 2012) for the same contrast metaphor > literal found the major clusters of activation in the left inferior frontal gyrus, and middle temporal gyrus (BA 45 and 21, respectively). The inferior frontal gyrus was found to be active bilaterally, although a left dominance was present. The other significant clusters were the left medial frontal gyrus (BA 9), inferior temporal gyrus (BA 20), and superior frontal gyrus (BA 8). With respect to metonymy, the only fMRI study, to our knowledge, that investigated this type of figurative language was conducted by Rapp and colleagues in 2011. The major clusters of activation for metonymy > literal were found in the left middle temporal gyrus (BA 21), and in the inferior frontal gyrus (left BA 9, 45, and right BA 47, 51). Other clusters were found in the bilateral middle frontal gyrus (BA 10), and superior medial frontal gyrus (BA 9).

The data from these studies suggest a partial overlap in the brain architecture supporting the processing of metaphor and metonymy, especially in the frontotemporal network. However, the lack of a direct comparison leaves open the extent to which processing metonymic and metaphoric language is supported by the same brain regions.

Interestingly, the frontotemporal network, observed both for metaphor and metonymy, includes regions belonging to both the perisylvian language network and the Theory-of-Mind (ToM) network. As Hagoort (2013) points out, language is supported by a network of brain structures, and language processing involves regions in the temporal and parietal cortices (the angular gyrus, and regions involved in attentional processes), as well as frontal brain structures in the inferior frontal gyrus, the dorsolateral prefrontal cortex, and the anterior cingulate cortex. These regions dynamically interact with each other in other functional networks as well. The inferior frontal gyrus plays a role in cognitive processes requiring theory of mind, together with the anterior cingulate cortex, and more temporal clusters. The ToM network includes clusters extending bilaterally from the temporal pole to the superior temporal gyrus. Regions involved are the inferior frontal gyrus, the temporal parietal junction, the medial prefrontal cortex, the medial orbitofrontal cortex, and the anterior cingulate cortex (Molenberghs et al., 2016). Yet, from previous studies, it remains unclear whether the activation of the frontotemporal network during metaphor and/or metonymy comprehension could be driven by inferential processes or by enhanced linguistic processing. Thus, the question of whether a possible difference between metaphor and metonymy might lie on differential activations in the Theory-of-Mind and/or in the language networks still calls for an answer.

1.2. The present study

The present project aims to assess, in an exploratory fashion, i) the extent to which metaphor and metonymy processing are supported by the same neural architecture, and ii) the extent to which the perisylvian language network and/or the ToM network are involved in processing metaphor and metonymy. In an event-related functional MRI study, participants were auditorily presented with literal, metonymic, metaphoric, and nonsensical sentences. Each sentence was paired to a picture representing the referent. Accordingly, some context for a correct interpretation of nonliteral language is in order, since the speaker's meaning, as argued above, goes beyond the semantics of the utterance itself and thus pragmatic processes play a fundamental role in hearers' meaning selection. The choice of using a picture, rather than a sentence alone, to provide a supportive context stems from the constraints that written text would create in terms of selecting the relevant features of the referent. In other words, a picture of the referent was meant to leave more degrees of freedom to the extraction of those features of the referent that are relevant for the correct selection of the nonliteral meaning. Participants, thus, were able to infer the target of the metaphoric and metonymic allusions thanks to *their own* extraction of its relevant characteristics from the picture.

Based on the previous literature, we predicted frontotemporal activations for both metaphor and metonymy processing. However, our investigation on the degree of overlap between conditions was exploratory in nature, since this was the first fMRI study directly comparing metaphor to metonymy comprehension. We considered two hypotheses with respect to the frontotemporal activity. If this was driven from the inferential processes necessary for understanding figurative language, we hypothesized a high correlation between the ToM localizer results and the activations in the figurative task. On the other hand, if metaphor and metonymy comprehension relies more on enhanced activation of the language network, we expected a higher correlation between the language localizer results and the activation in the figurative task.

2. Method

2.1. Participants

Thirty-seven healthy right-handed native speakers of Dutch (22 females, 15 males, mean age = 23.64, SD = 5.48, range 18-46) participated in the study. All of the participants were recruited through the online Radboud Research Participation System and received payment for their participation in the study. Participants had normal or corrected-to-normal vision, and they reported no dyslexia or other reading and hearing problems, no colorblindness, and no history of neurological and psychiatric disease. All participants filled in the informed consent and screening form, and they received instructions about the tasks on paper before the start of the experiment. Data from three participants were discarded due to motion artifacts. Hence, data from 34 participants were included in the analyses (21 females, 13 males, mean age = 23.08, SD = 4.13, range 18-36). The number of subjects included was in line with the results of the power analysis (α *err prob* = .05; *Power (1-β err prob)* = .80) we ran based on data from Pinango et al. (2016) for the contrast conventional vs systematic metonymies.

2.2. Material and design

The study consisted of three tasks: the figurative language task and two localizers, for language (Lam et al., 2016) and Theory-of-Mind (Dodell-Feder et al., 2011), respectively.

2.2.1. Figurative Language Task

The experimental material of the main task consisted of 160 spoken sentences and 40 pictures across 4 conditions. Each picture was paired with 4 sentences, one per condition. These sentences were the same except for the target word, which could be used in a literal, metonymic, metaphoric, or nonsensical way with respect to the referent represented in the picture. An example stimulus set is presented in *Figure 1*. The nonsensical condition was introduced in order to control for possible brain activations elicited by prediction error effects. Accordingly, target words that were not literal with respect to the referent represented in the picture had a lower predictability than literal ones, and were thus likely to elicit brain activations related to unexpectancy, rather than figurativity.

The experiment consisted of 160 events, i.e. picture-sentence pair presentation, in an event-related design. In order to ensure that participants were always attending to both sentences and pictures, 10% catch trials (i.e. 16 events) were additionally presented during the experiment, for a total of 176 trials. These were constituted by one repetition of the same stimulus pair that participants were instructed to detect via a button press.

Pictures were taken from online sources via Google and selected on the basis of a referent pre-test, in which fourteen additional participants from the same participant pool, who did not take part in the main experiment performed an online picture-naming task. The pre-test sample was comprised of 10 females, and 4 males, all university students. The test was anonymous, hence we did not collect any additional information from the pre-test sample. Pictures were selected according to a minimum 90% agreement with respect to label consistency. They were subsequently resized via the online software Bulk Resize (Wruff Pty Ltd., https://bulkresizephotos.com/en/copyright), and additionally scaled in Presentation to fit the center of the screen and participants' foveal vision.

The sentences were recorded in a sound proof booth and spoken by a female native speaker of Dutch. They were digitized at a sample frequency of 44.1 kHz, and equalized in maximum amplitude in Praat software (version 5.3.55, Boersma & Weenink, 2019). The volume was normalized in Matlab after recording. Mean duration of the sentences was 2732 ms (SD = 418), matched across conditions (see *Table 1*).

Target words in the metonymic, metaphoric, and nonsensical conditions were controlled for semantic distance with respect to the referent, i.e. the literal counterpart (Snaut, Mandera et al., 2017). In line with the definition of metaphor and metonymy, a semantic analysis showed greater distance from the literal counterpart for metaphor than for metonymy (*Table 2*). Target words were also controlled for length and frequency (SUBTLEX-NL online database, Keuleers, et al. (2010), see *Table 3* and *Table 4*).

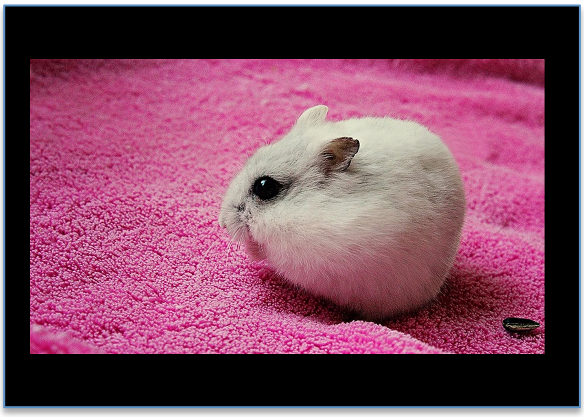

Figure 1. Example of picture-sentence stimulus pair.

Literal: De *hamster* speelt in zijn kooi (Eng: The *hamster* is playing in its cage)

Metonymy: De *snuit* speelt in zijn kooi (Eng: The *snout* is playing in its cage)

Metaphor: De *sneeuwbal* speelt in zijn kooi (Eng: The *snowball* is playing in its cage)

Nonsensical: De *komkommer* speelt in zijn kooi. (Eng: The *cucumber* is playing in its cage)

Table 1. Results of separate *t* tests comparing the mean duration of the sentences across different conditions.

T tests	p-values
Literal and metonymy	0.8382
Literal and metaphor	0.7165
Literal and nonsensical	0.8637
Metonymy and metaphor	0.9031
Nonsensical and metonymy	0.9598
Nonsensical and metaphor	0.8437

Table 2. Semantic distance between target words from literal, metonymic, metaphoric and nonsensical condition, and relative *t* tests. LIT=literal; MN=metonymy; MP=metaphor; NS=nonsensical.

	LIT-MN	*LIT-MP*	*LIT-NS*
AVG	0.6389	0.8089	0.8646
SD	0.1300	0.1148	0.0815

T tests (*p* values)	
Lit-Mn vs Lit-Mp	1.7187e-07
Lit-Mn vs Lit-Ns	1.9212e-13
Lit-Mp vs Lit-Ns	0.0038

Table 3. Word length and word frequency per condition as taken from the SUBTLEX-NL database.

	Length (mean and SD)		Frequency (mean and SD)	
Literal	7	2.23	34.76	79.78
Metonymy	6.45	2.37	31.16	52.08
Metaphor	6.92	2.35	27.10	50
Nonsensical	6.45	2.38	24.97	39.12

Table 4. Outcome of separate *t* tests comparing word length and word frequency across conditions.

T tests	Length *p* values	Frequency *p* values
Literal and metonymy	0.2890	0.8142
Literal and metaphor	0.8820	0.6147
Literal and nonsensical	0.2900	0.4935
Metonymy and metaphor	0.3770	0.7284

Nonsensical and metonymy	1	0.5533
Nonsensical and metaphor	0.3780	0.8351

2.2.2. Language Localizer Task

The localizer task for language was identical to the one described in Lam et al. (2016). It consisted of 12 blocks of stimuli across two conditions. Each block was constituted of either five sentences or five word-lists, depending on the condition. Streams of 9-to-15 words were sequentially presented as random words in the word-list condition and as meaningful sentences in the sentence condition. The word-list condition differed from the sentence condition for the overall lack of a coherent meaning. Within each block, and across conditions, a number of nonwords were presented, and participants had to detect these via button press. The blocks for each condition were presented in alternation and in a counterbalanced fashion across participants. (For more details, see Lam et al., 2016).

2.2.3. Theory-of-Mind Localizer Task

The task for localizing the core brain regions underlying ToM processing was meant to target top-down inferential ToM processes (for a meta-analysis and discussion, see Molenberghs, et al., 2016). Twenty short stories followed by a true/false question concerning the same story were presented in random order across two conditions (an example is shown in *Fig.2*). In both conditions participants had to represent false contents. However, the contents to be represented differed. In the False Belief condition, the story concerned false beliefs, as opposed to the False Photograph condition, in which the stories were descriptions of outdated (hence false) photographs or maps. (For more details, see Dodell-Feder et al., 2011).

2.3. Procedure

After receiving instructions on the tasks, participants had some time for practicing in the scanner before the experiment start. The order of the tasks to practice was the same as in the real experiment. Each participant started with a practice session for the main figurative language task, followed by a practice for the language, and finally for the ToM localizer task. The stimuli used for practicing the figurative language task comprised of one block of 32 picture-sentence pairs that were not subsequently used during the real task (4 sentences per picture across conditions, for a total number of pictures equal to eight). For the language localizer practice task, 10 stimuli were presented, five per condition. These were different from those used in the real task. Finally, for the ToM practice, one of the stimuli from the real task was presented.

After the practice, an additional sound check was performed, in order to adjust the volume of the spoken items while the scanner was on. Participants were asked to give feedback via button press on how well they could hear spoken items that were presented binaurally through the same special MRI-compatible headphones that were used during the real experiment, until the spoken items were clearly audible.

After the volume check, participants started with the main task. The 160 events plus 10% catch trials were presented one after the other and across four blocks of 40 stimuli pairs each. Blocks were separated by a break (i.e. the Dutch word *pauze,* “pause” in English, was presented on screen), and participants were asked to press the right index finger whenever they felt ready to continue. Each trial consisted of a fixation cross, jittered in time (between 2000 and 6000 ms), followed by a picture-sentence pair. First, a picture was presented in the center of the screen. After 50 ms, a sentence was binaurally presented through the headphones. An example of an experimental trial is presented in *Figure 2*. To ensure participants were attending both the visual and auditory stimuli, they were asked to press the right index finger every time the same stimulus, i.e. same picture-sentence pair, was presented on two subsequent trials.

Stimulus presentation was counterbalanced across participants. Thirty-seven different lists, one per participant, were created in Mix (van Casteren & Davis, 2006) allowing for the randomization of the stimuli. Since each picture was presented 4 times, once per condition, we constrained the randomization using the following parameters. The same picture was presented once per block, and with a minimum 10-item distance from its next presentation. Stimuli belonging to the same condition were additionally spaced out via a minimum distance of 10 items across presentations.

The task was programmed in and presented via Presentation software (version 16.4, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

After the main task, each participant performed the language localizer task followed by the ToM localizer task. For the language localizer task, the instructions were given by the experimenter through the microphone from the control room. Each trial started with a fixation cross that was presented in the center of the screen. This was followed by the name of the condition and the number of the block (e.g. sentence/01). A word-by-word list was then presented on the screen as a coherent sentence in the sentence condition, or as a random list in the word-list condition. Participants were asked to press the right index finger, whenever they detected a non-word. Trials were randomized and counterbalanced across participants.

For the ToM localizer task, the instructions were presented on the screen. Each trial consisted of a fixation cross (jittered in time between 6000 and 8000 ms) and a story-question pair. Each story was presented visually for 10 seconds, after which the related true/false question appeared on the screen. Participant were asked to press the right or left index finger for answering. There were a total of 20 trials, 10 per condition. Trials were randomized and counterbalanced in Presentation.

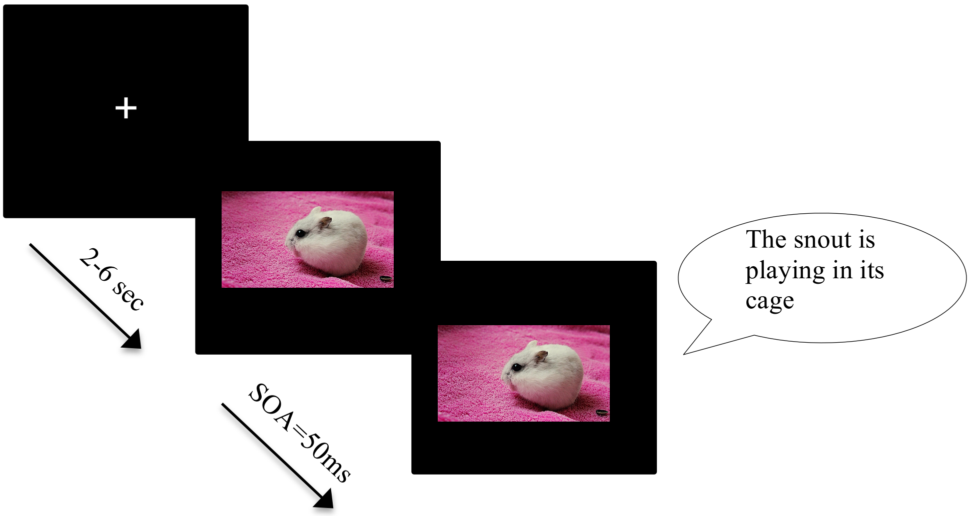

Figure 2. Example of one experimental trial for the figurative language task.

2.4. fMRI data acquisition

Data were acquired from a 32-channel head coil in a Siemens 3T Prisma MRI scanner. The same GRE-EPI, 2mm isotropic sequence was used for all of the tasks, as it allows for good resolution in frontal and temporal brain structures. T2*-weighted functional scans were acquired in the transverse plane with 66 slices per volume (TR = 1000 ms, TE = 34 ms, flip angle = 60°, FoV = 210 mm, voxel size = 2x2x2 mm, multiband accelerator factor = 6).

There were 3 functional runs. The first one for the main task lasted around 33 minutes, and was followed by a second 12-minute and a final 8-minute run for the language and ToM localizer, respectively. After the final functional run, a high-resolution anatomical scan with full coverage of the brain was acquired in the sagittal plane via a 3D pulse sequence weighted for T1 contrast (MPRAGE, TR = 2.3 s, TE = 3.03 ms, flip angle = 8°, FoV = 256 mm, voxel size = 1x1x1 mm, 192 slices GRAPPA acceleration factor = 2).

2.5. fMRI data analysis

Data were preprocessed in SPM software package (Statistical Parametric Mapping, version 12, www.fil.ion.ucl.ac.uk/spm/) via its implementation in Matlab (version 2019a, Mathworks Inc., Sherborn, MA, USA). Functional scans were all subjected to the same preprocessing pipeline. A Voxel Displacements Map (VDM) was created via the FieldMap toolbox and used for unwarping the geometrically distorted EPI images. The field map was co-registered to the first functional volume and used during the within-subject registration and unwarping of the time series for increasing the quality of motion artifacts removal. Functional scans were then realigned to the first volume, co-registered to the anatomical image, normalized to the MNI (Montreal Neurological Institute) space (EPI template), and smoothed (6 mm FWHM Gaussian for noise reduction in single subject).

2.5.1. Statistical modeling: figurative language task

At the single-subject level, statistical maps were created based on the General Linear Model (GLM). Auditory events were included in the analysis as events of interest. Each experimental condition consisted of 40 events (plus 16 events for the catch trials, which were also modeled for a motor activation sanity check at the individual level, but not further taken into account). Event-related responses were convoluted with the canonical Hemodynamic Response Function (HRF), and modeled specifying their onset and duration times. Additional regressors were the 6 movement parameters estimated during the realignment preprocessing step and the temporal derivative of the HRF. Data were high-pass filtered at 128 s cutoff. T-contrasts were specified at the whole-brain level for each condition against implicit baseline. Additional contrasts of interest were entered at the single-subject level for each condition > literal, metaphor > metonymy, and metonymy > metaphor. The contrasts of interests are listed in *Table 5*.

At the second level, contrasts images were tested via one sample t tests. All the activation maps at the second level were double-checked by running paired t tests from the single-subject condition > implicit baseline for the same contrasts. Whole-brain activation maps were thresholded at a voxel-wise uncorrected $p <= .001$, and subsequently assessed for cluster-wise significance in the context of FWE ($p <= .05$). For the subtler contrasts metaphor > metonymy and metonymy > metaphor, activations maps were further inspected at an uncorrected p value of .01, in order to avoid Type II error.

Table 5. List of contrasts of interest at first- and second-level analysis

Contrasts of interest
Nonsensical > Literal
Metonymy > Literal
Metaphor > Literal
Metonymy > Metaphor
Metaphor > Metonymy

2.5.2. Statistical modeling: localizer tasks

The same first-level pipeline that was applied to the data from the figurative language task, was also used for the localizer tasks.

2.5.2.1. Language localizer task

The language localizer task was modeled as a block design consisting of 12 blocks across two conditions. Individual statistical maps at the single-subject level were obtained in the context of the General Linear Model. Block-related responses were convoluted with the canonical HRF, and modeled in terms of onset and duration times. The temporal derivative of the HRF was not modeled in this context as this was a block design. The 6 movement parameters were added to the design matrix as additional regressors. Data were high-pass filtered at a 128 s cutoff. The contrasts entered at the first-level were the contrast image of interest sentence > word-list (i.e. activation map in the sentence condition versus that in the word-list condition), and one contrast image per condition against implicit baseline. At the group-level, the activation map for the contrast image sentence > word-list was inspected at an uncorrected voxel-wise threshold of p = .001, and corrected for multiple comparison in the context of FWE ($p <= .05$).

2.5.2.2. ToM localizer task

The ToM localizer task was modeled as an event-related design, in which the events of interest were the 20 stories presented across the False Belief and the False Photograph condition. Event-related responses were convoluted with the canonical HRF and its temporal derivative, and modeled in terms of onset and duration times. Additional regressors at the individual level were the 6 motion parameters from the preprocessing step. The contrast images entered at the first level were each condition against baseline, and the contrast of interest False Belief versus False Photograph. This contrast image was further investigated at the group level and inspected at a *p* value of .001. Multiple comparison correction was applied as for the other tasks in the context of FWE correction.

Before performing the group-level analysis, all the functional data across runs were checked for the presence of outliers that survived preprocessing. Spikes were removed via ArtRepair software (version 5b, Mazaika, Hoeft, Glover et al., 2009), which is implemented in Matlab.

3. Results

3.1. Figurative task

3.1.1. Behavioral

The accuracy rate for catch trials detection in the figurative task revealed that participants were paying attention to the stimuli. All of the participants reported at least 81% accuracy in catch trials detection (M = 15.70 of correct answers across 16 trials, SD=0.70).

3.1.2. Whole-brain analysis

Since the study was exploratory in nature, we performed a whole-brain analysis for each contrast of interest. In particular, we were interested in any difference between the figurative conditions and the literal condition, and between metaphor and metonymy. In addition, we exploited the contrast nonsensical versus literal as a control condition in order to rule out activations elicited by (violated) expectancy processes rather than figurativity. The results for each contrast are displayed in *Table 6*.

The *t* tests we ran at the second level for metonymy vs literal and metaphor vs literal both revealed increased frontotemporal network activation. The contrast nonsensical versus literal revealed a similar activity pattern. The activity clusters spanned across both hemispheres, although a left lateralization was present (*Figure 3A*).

The major clusters of activation for metonymy vs literal were found in the left hemisphere (k = 7563), in particular in the middle and superior temporal gyri (BA 21, 22), and inferior frontal gyrus (BA 45). The homologous areas in the right hemisphere included the superior temporal gyrus, the superior temporal pole (k = 1363), and the inferior frontal gyrus (k = 387). Positive and negative activity clusters for the contrast metonymy vs literal are depicted in *Figure 3B*.

The major activity clusters for the contrast metaphor vs literal were found in the left hemisphere, specifically in the superior and middle portions of the temporal gyrus (BA, 22, 21, k = 4517), up to the fusiform gyrus (BA 37), and in the left inferior frontal gyrus (BA 45, 47, k = 1849). Increased activity was also found more occipitally in the left hemisphere, and in the temporal gyrus in the right hemisphere. In addition, a frontoparietal network of regions, including the bilateral anterior and middle cingulate gyrus, the right angular gyrus and medial superior frontal gyrus showed a decrease in activity for the same contrast. Positive and negative activity clusters for the contrast metaphor vs literal are depicted in *Figure 3C*.

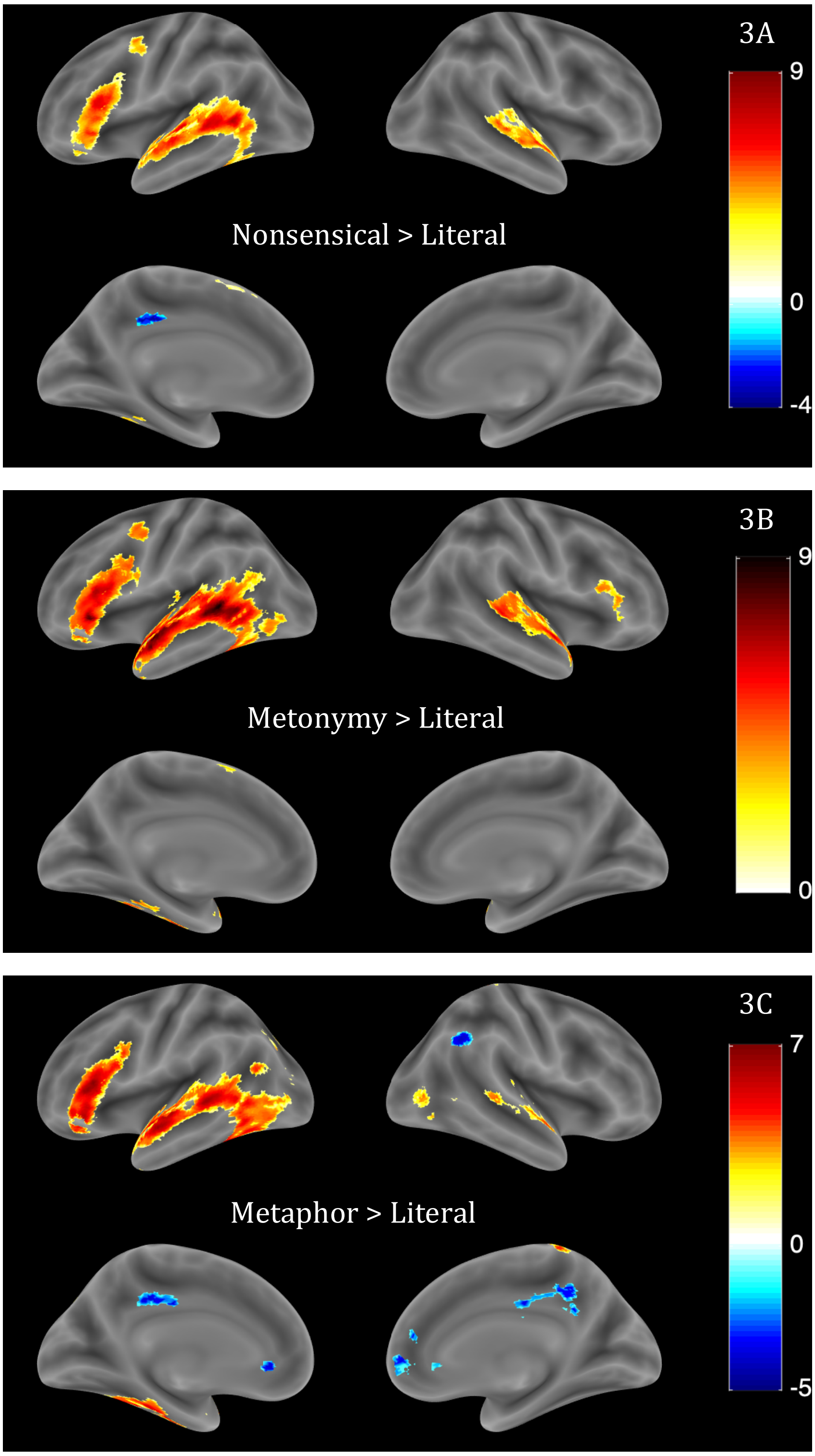

Figure 3. Lateral and sagittal view (top and bottom row, respectively) of the left and right hemisphere.
Positive (red) and negative (blue) activity clusters for the contrast Nonsensical vs Literal (3A), Metonymy vs Literal (3B), and Metaphor vs Literal (3C).

All contrasts show increased left-lateralized frontotemporal activity with a RH contribution at the level fo the superior temporal gyrus, and superior temporal pole (A, B, C).
(A) The middle cingulate gyrus (blue) showed a decrease in activity for the contrast Nonsensical > Literal.
(C) The contrast Metaphor vs Literal shows increased frontotemporal and decreased frontoparietal activity.

Cluster-wise correction *p(FWE)* 0.05

The middle cingulate gyrus was also less active when we inspected the contrast nonsensical vs literal. Nonsensical sentences elicited the major cluster of activity in the left hemisphere (middle temporal gyrus (k =2704), and inferior frontal gyrus (BA 45, k = 1792). In the right hemisphere, the right superior temporal gyrus was found to be significantly active, extending until the right superior temporal pole (BA 38).

For the contrasts metonymy vs metaphor, the only cluster surviving FWE correction was found more posteriorly in the brain, and included the bilateral precuneus (BA 7), with extensions to the left posterior cingulate gyrus (BA 23, 31). These regions, in particular, showed to be modulated by the type of figurativity being processed (i.e. metaphor or metonymy), showing an increase in activity during metonymy processing (vs metaphor), and a decrease during metaphor comprehension (vs metonymy). For the contrast metaphor > metonymy, no suprathreshold positive cluster was found at $p <= .001$. Results were further investigated at a lower threshold at a voxel-wise $p <= .01$, without, however, leading to any additional finding.

Table 6. Results of the main figurative task. Shows all local maxima separated by more than 8 mm. Regions were automatically labeled using the AAL2 atlas. x y and z =Montreal Neurological Institute (MNI) coordinates in the left-right anterior-posterior and inferior-superior dimensions respectively. Thresholding t > 3.3563

Contrast Name				MNI Coordinates		
	Region Label	Extent	t-value	x	y	z
nonsensical > literal	Temporal_Mid_L	2704	9.5662	-50	-52	2
	Temporal_Mid_L	2704	7.7704	-56	-38	2
	Temporal_Mid_L	2704	7.7289	-58	-14	-2
	Temporal_Sup_R	871	7.7541	54	-30	4
	Temporal_Sup_R	871	6.6865	62	-6	-4
	Temporal_Pole_Sup_R	871	6.3634	60	4	-8
	Frontal_Inf_Tri_L	1792	7.5008	-38	28	-2
	Frontal_Inf_Tri_L	1792	7.4720	-48	20	22
	Frontal_Inf_Tri_L	1792	7.2542	-44	14	28
	Frontal_Mid_2_L	203	5.3370	-34	0	50
	Precentral_L	203	4.6179	-48	0	50
	Location not in atlas	203	4.4384	-32	0	40
	Supp_Motor_Area_L	303	5.2999	-2	6	60
	Frontal_Sup_Medial_L	303	4.2839	-2	26	56
	Cingulate_Mid_L	169	-4.5133	0	-34	40
	Cingulate_Mid_L	169	-4.3515	-8	-32	34
metonymy > literal	Temporal_Sup_L	7563	9.1753	-56	0	-8
	Temporal_Mid_L	7563	9.1339	-52	-42	4
	Frontal_Inf_Tri_L	7563	8.6995	-42	30	-2
	Temporal_Sup_R	1363	6.7791	60	2	-10
	Temporal_Sup_R	1363	6.6233	46	-34	4
	Temporal_Pole_Sup_R	1363	6.4393	50	16	-22

	Frontal_Inf_Tri_R	387	5.0207	56	34	6
	Frontal_Inf_Tri_R	387	4.8615	54	22	22
	Frontal_Inf_Tri_R	387	4.5889	52	32	16
	Precentral_L	200	4.7991	-36	2	48
	Precentral_L	200	4.6601	-44	-2	46
	Supp_Motor_Area_L	174	4.7476	-2	8	60
	Supp_Motor_Area_L	174	4.3260	-4	4	68
	Frontal_Sup_2_L	174	3.6681	-14	6	54
metaphor > literal	Frontal_Inf_Tri_L	1849	7.6361	-48	26	10
	Frontal_Inf_Orb_2_L	1849	6.7792	-44	32	-8
	Frontal_Inf_Tri_L	1849	6.3221	-54	24	20
	Temporal_Sup_L	4517	7.3200	-54	-8	-10
	Temporal_Mid_L	4517	7.3140	-54	-38	2
	Fusiform_L	4517	6.9171	-38	-38	-22
	Postcentral_R	154	5.1676	12	-40	74
	Precentral_R	154	4.1123	18	-28	76
	Temporal_Pole_Sup_R	149	5.1490	60	4	-6
	Temporal_Sup_R	149	4.4823	64	-8	-2
	Occipital_Sup_L	169	4.9085	-24	-72	38
	Occipital_Mid_L	169	4.2120	-30	-74	30
	Occipital_Mid_L	169	3.8832	-28	-82	34
	Temporal_Sup_R	163	4.7404	46	-34	8
	Temporal_Sup_R	163	4.3974	54	-32	4
	Temporal_Sup_R	163	4.2033	50	-22	0
	Occipital_Mid_R	211	4.5017	48	-78	6
	Temporal_Inf_R	211	4.4185	52	-70	-4
	Temporal_Mid_R	211	3.7239	58	-62	4
	Cingulate_Mid_L	613	-5.6206	0	-32	40
	Cingulate_Mid_R	613	-5.2595	6	-50	36
	Cingulate_Mid_L	613	-4.7549	-2	-20	36
	Frontal_Sup_Medial_R	532	-5.2225	4	52	0
	Cingulate_Ant_L	532	-4.8095	-4	32	-6
	Cingulate_Ant_R	532	-4.6476	12	42	14
	Angular_R	161	-4.6584	50	-56	40
metonymy > metaphor	Precuneus_R	421	5.9659	4	-62	32
	Cingulate_Post_L	421	4.6595	-12	-48	30
	Precuneus_L	421	4.4049	-8	-58	30
metaphor > metonymy	Precuneus_R	327	-5.6720	4	-62	32
	Cingulate_Post_L	327	-4.5550	-12	-48	30
	Precuneus_L	327	-4.3616	-8	-56	30

3.2. Localizer tasks

3.2.1. Whole-brain analysis

The results of the localizer tasks are depicted in *Figure 4* and listed in *Table 7* and *8*.

The contrast Sentences > Word-list in the Language localizer showed a left lateralized network of brain regions, including the left middle temporal pole, and the superior and middle portions of the temporal gyrus (*Figure 4A*).

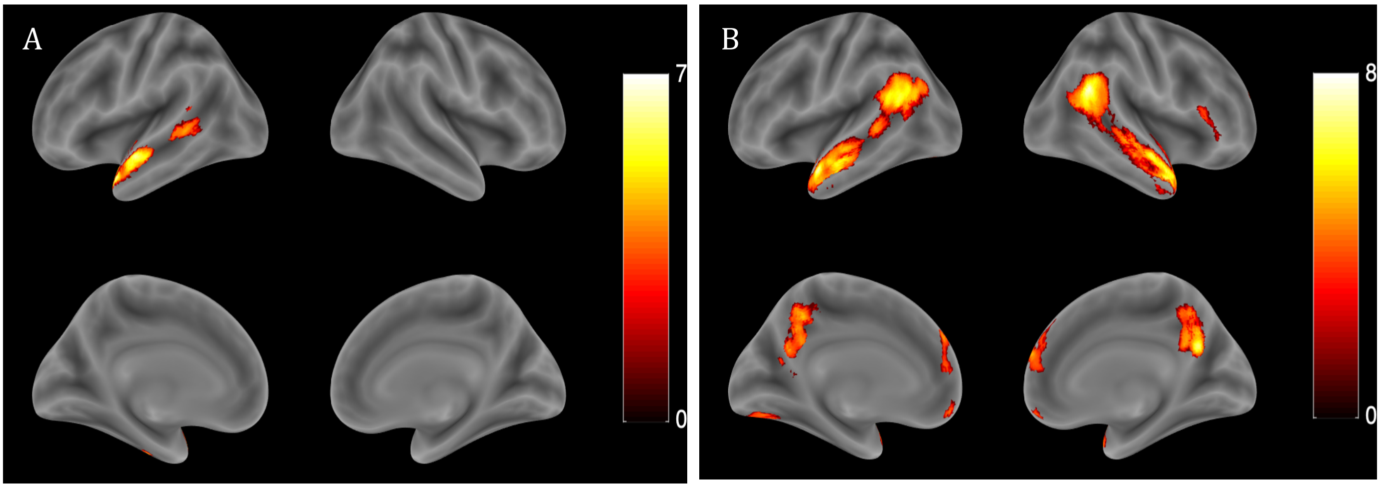

Figure 4. L/R medial/lateral view of *(A)* Language localizer intensity image of the thresholded whole-brain map (FWE p .05) for the contrast sentence > word-list; *(B)* ToM localizer intensity image of the thresholded whole-brain map (FWE p .05) for the contrast False Beliefs > False Photograph.

With respect to the localizer task for ToM, for the contrast False Beliefs > False Photographs, we found increased activity in a network of brain areas across both hemispheres (*Figure 4B*). This included the angular gyrus, temporal gyrus (middle portion), precuneus, medial superior frontal cortex, prefrontal cortex, and inferior frontal gyrus (k = 3423 in left hemisphere; k = 3320 in right hemisphere).

Table 7. Language localizer – shows all local maxima separated by more than 8 mm. Regions were automatically labeled using the AAL2 atlas. x y and z =Montreal Neurological Institute (MNI) coordinates in the left-right anterior-posterior and inferior-superior dimensions respectively. Thresholding t > 3.3563

Contrast Name				MNI Coordinates		
	Region Label	Extent	t-value	x	y	z
Sentence > Word-list	Temporal_Pole_Mid_L	843	7.8316	-48	16	-26
	Temporal_Sup_L	843	7.1276	-54	-8	-10
	Temporal_Mid_L	843	6.3556	-50	-6	-18
	Temporal_Mid_L	363	4.9326	-48	-44	0
	Temporal_Mid_L	363	4.5138	-56	-52	8
	Temporal_Mid_L	363	4.4285	-52	-34	2

Table 8. ToM localizer – shows all local maxima separated by more than 8 mm. Regions were automatically labeled using the AAL2 atlas. x y and z =Montreal Neurological Institute (MNI) coordinates in the left-right anterior-posterior and inferior-superior dimensions respectively. Thresholding t > 3.3479

Contrast Name				MNI Coordinates		
	Region Label	Extent	t-value	x	y	z
False Beliefs > False Photographs	Angular_R	3423	8.7665	48	-52	24
	Temporal_Mid_R	3423	8.7350	54	4	-24
	Temporal_Mid_R	3423	8.2123	54	-12	-14
	Temporal_Mid_L	3320	8.6266	-52	2	-20
	Temporal_Mid_L	3320	8.1097	-58	-12	-16
	Temporal_Mid_L	3320	7.5756	-44	-52	18
	Precuneus_R	1585	8.4442	2	-60	22
	Precuneus_L	1585	6.7439	0	-50	42
	Precuneus_R	1585	6.7419	4	-56	36
	Frontal_Sup_Medial_R	930	5.9974	10	60	24
	Frontal_Sup_Medial_L	930	5.2097	-8	52	36
	Frontal_Sup_Medial_L	930	5.1941	2	50	38
	Frontal_Med_Orb_L	196	5.8147	-4	56	-12
	Frontal_Med_Orb_R	196	5.6608	4	56	-14
	Lingual_L	247	4.7273	-24	-74	-4
	Fusiform_L	247	4.6985	-30	-70	-10
	Frontal_Inf_Tri_R	151	4.7205	58	26	18
	Frontal_Inf_Orb_2_R	151	4.3668	52	30	-4
	Frontal_Inf_Tri_R	151	4.3315	56	26	2

3.3. Overlays between localizers and figurative task

The extent of overlap between the regions active for ToM and those active for language processing are shown in *Figure 5*. This in particular concerns the middle temporal gyrus and temporal pole in the left hemisphere.

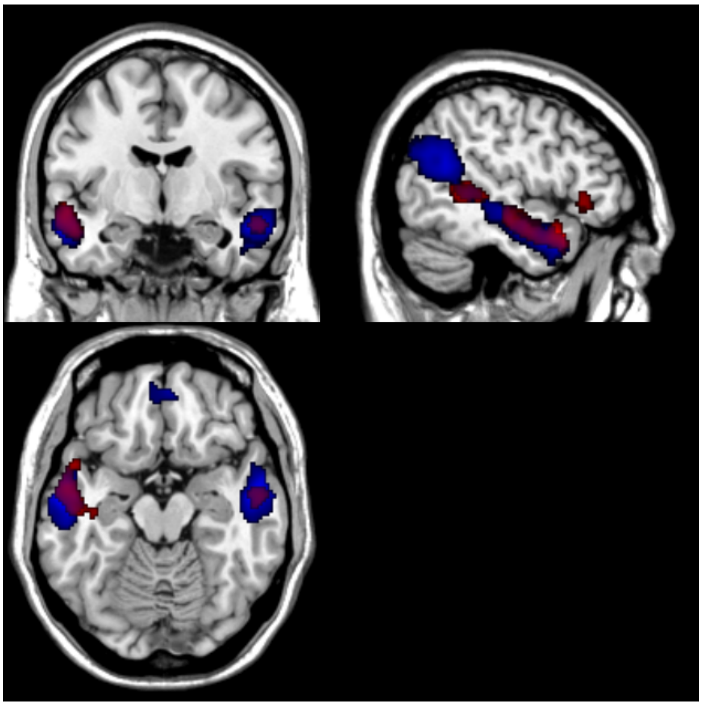

Figure 5. Degree of overlap (in purple) between the localizer for language (red) and for ToM (blue). Coronal, sagittal and axial view (from left to right). The overlap extends from the middle temporal gyrus to the temporal pole in the left hemisphere.

Figure 6 and *Figure 7* show the overlap between the nonliteral vs literal contrasts and each of the localizers. Both contrasts, metonymy vs literal and metaphor vs literal show similar degree of overlap with respect to each of the localizer tasks, in particular in the right angular gyrus and in the left middle temporal gyrus.

Figure 8 depicts the overlay between the ToM localizer activity clusters and the contrast metonymy vs metaphor, and shows an almost perfect overlap in the bilateral precuneus and cingulum (peak coordinates for ToM cluster, 2 -60 22; peak coordinates for metonymy > metaphor cluster, 4 -62 32).

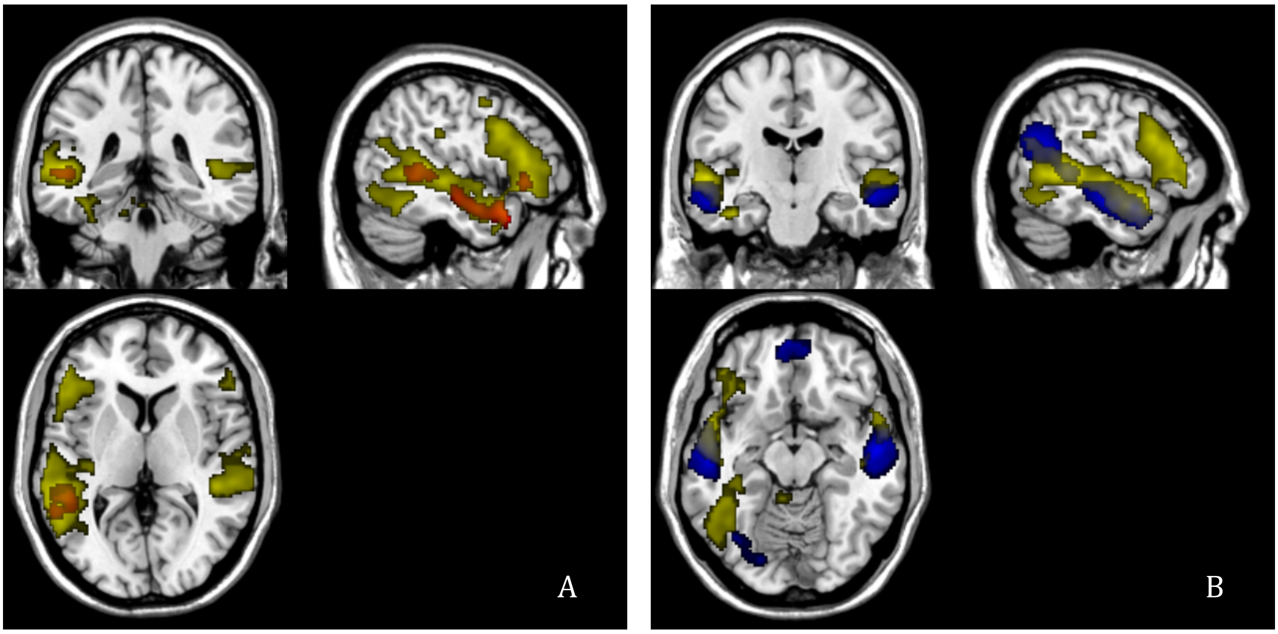

Figure 6. (A) Overlay between the regions active for the language localizer (red) and those active for the contrast metonymy vs literal (yellow). *(B)* Overlay between the regions active for the ToM localizer (blue) and those active for the contrast metonymy vs literal (yellow).
Coronal, sagittal, and axial view (from left to right). Left hemisphere of the left, right hemisphere on the right. The sagittal view depicts the left hemisphere.

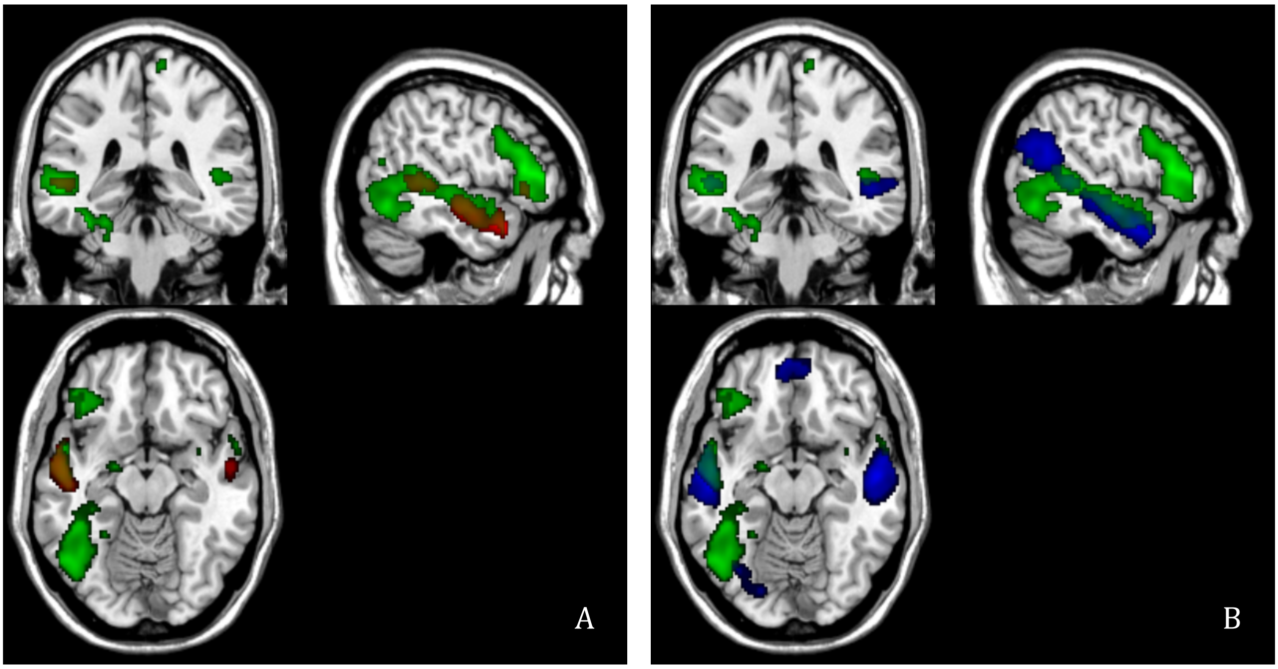

Figure 7. (A) Overlay between the regions active for the language localizer (red) and those active for the contrast metaphor vs literal (green). (B) Overlay between the regions active for the ToM localizer (blue) and those active for the contrast metaphor vs literal (green).
Coronal, sagittal, and axial view (from left to right). Left hemisphere of the left, right hemisphere on the right. The sagittal view depicts the left hemisphere.

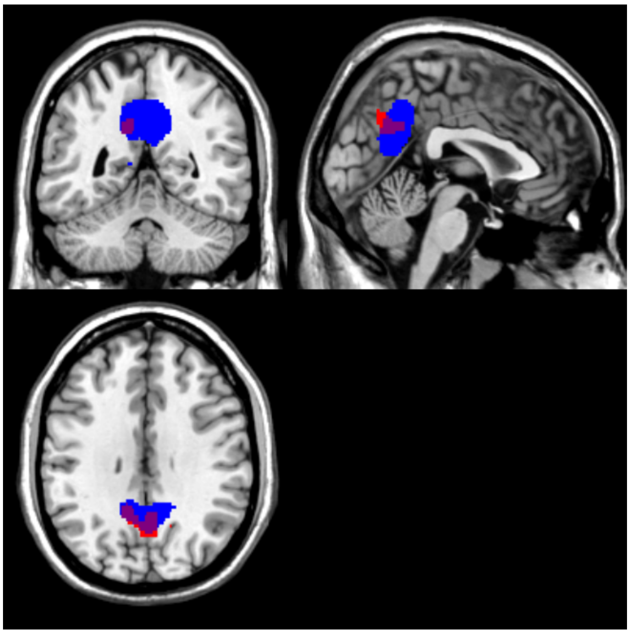

Figure 8. Overlap (purple) in the precuneus and cingulum, bilaterally. The blue area depicts the activation during the ToM task, whereas the red that for the contrast metonymy > metaphor.

4. Discussion

The present study investigated the neural basis of metonymy and metaphor comprehension, and the degree of involvement of the language and the ToM network while processing these two types of figurative language. In particular, we were interested in the investigation of whether the brain architectures underlying metaphor and metonymy comprehension differed depending on the type of figurative language being processed, and whether these brain regions differentially mirrored enhanced linguistic processing or inferential processes.

4.1. Knowledge integration in the frontotemporal semantic network

In line with the literature (Rapp et al., 2011; Rapp et al., 2012; Bohr et al., 2012), processing metaphoric and metonymic sentences (versus literal) elicited increased frontotemporal activity, including regions in the left inferior frontal gyrus, and left middle temporal gyrus (middle and superior portions). In the right hemisphere, common regions involved in metonymy and metaphor processing were found in the right superior temporal gyrus with extension up to the superior temporal pole. These areas were also found to be upregulated during the processing of nonsensical sentences. The consistency with which the activity in these areas was upregulated during the processing of nonsensical, metaphorical and metonymic sentences, suggests that their recruitment might not be specific to figurative language comprehension. Accordingly, the middle temporal gyrus, together with the inferior frontal gyrus, and the superior temporal gyrus have been shown to be part of the core semantic network (e.g. Hagoort, 2005; Menenti et al., 2009), hence to be key regions during semantic processing (Jackson et al., 2016). In addition, the temporal lobe, including the right superior temporal gyrus has been found to be involved in meaning construction (Stolk et al., 2014). Moreover, the same regions have been implicated in inferential processes in a study investigating how conversational implicatures are resolved at the neural level (Jang et al., 2013). In this study, the left inferior frontal gyrus was additionally modulated by the degree of implicitness of the stimuli. Accordingly, this region has been long recognized to be involved in conflict resolution (e.g. Novick, et al., 2009; Thothathiri, et al., 2012) when different representations compete at the lexical, syntactic, semantic and/or conceptual level, like in the present study.

A common feature of the nonliteral conditions (figurative, and nonsensical alike) was the semantic distance from the literal counterpart (see *Table 2*). Participants, while processing target words that were not literal with respect to the referent represented in the picture, were thus likely to experience a gap between what they were expecting to hear depending on the picture representing the referent, and what they actually heard (i.e. target words that were nonliteral with respect to the referent). This, we argue, independently on whether the target word was nonsensical, metaphorical, or metonymic, and as long as there was a discrepancy between the literal meaning of the target word and the picture representing the referent. Our hypothesis is that the gap between the literal meaning of the target words, and the referent represented in the picture created a representational conflict that was in turn detected by the left inferior frontal gyrus and solved in terms of world knowledge integration (for a discussion, see Hagoort et al., 2004; Hagoort, 2005). General world knowledge anomalies have been shown to modulate activity in the inferior frontal gyrus in studies investigating world knowledge integrative processes during sentences comprehension (Hagoort et al., 2004; Hagoort, 2005). For comprehending, for example the metaphoric reference to a white hamster in the sentence 'The snowball is playing in its cage' (*Figure 1*), it is necessary to access world knowledge about snowballs (i.e. snowballs are white and round), and integrate this with what we know about hamsters (i.e. they can be white, fluffy, and round-shaped). Such an integration process might thus be a possible strategy for extracting the relevant features of the referent and those of the nonliteral word, and mapping them together for comprehending the figure of speech. The left inferior frontal gyrus would then access information that is stored in temporal regions (Indefrey & Cutler, 2004), and map them together into a single representation (Hagoort, 2005). Accordingly, figurative language relies on the coupling of two competing representations for the same word, one that is tied to its semantics (and thus extractable thanks to our knowledge about the world, e.g. snowballs are white), and the other that is derivable from it (via knowledge integration). Although this does not fully apply to our nonsensical condition, in which the target word was completely unrelated and without any sense with respect to the context provided by the picture and the sentence itself, it is still possible that participants tried to make sense of it. Thus, processing a representational conflict similar to the one elicited by metaphors and metonymies, and trying to find a possible association between the nonsensical word and the referent in the picture.

4.2. Representational conflict in metaphor and metonymy processing

4.2.1. The right inferior frontal gyrus in metonymy processing

Activity in the right inferior frontal gyrus was found to be upregulated during metonymy processing. In general, metonymy seemed to elicit activation in the frontotemporal network that comprised the right hemisphere to a greater extent than metaphor. The upregulation of the right inferior frontal gyrus during metonymy processing is in line with the neuroimaging literature (Rapp et al. 2011), and has been explained in terms of additional cognitive effort for integrating the nonliteral reference into the overall meaning of the sentence in which the metonymy is embedded. In line with this view, the right inferior frontal gyrus contribution has been reported in numerous studies investigating different types of figures of speech (for a review, see Rapp et al., 2012). In the present study, however, the right inferior frontal gyrus was upregulated during metonymy but not during metaphor processing, suggesting that the type of figurativity being processed modulated the activity in this brain region.

This finding fits well with the present debate on the reasons driving the right hemisphere upregulation during figurative language processing. The right hemisphere activation during metaphor processing has been hypothesized to be (inversely) modulated by factors such as context, and familiarity of the stimuli, and not to be specific to figurative language processing (Diaz & Eppes, 2018). Yet, some fMRI studies reported activity in the right inferior frontal gyrus during the processing of novel metaphors (e.g. Yang et al., 2010), while others reported its upregulation during conventional metaphors comprehension (e.g. Ahrens et al., 2007). This inconsistency leaves room to the role that familiarity might play in the right hemisphere involvement during metaphor processing, and cannot explain the differential activity we found either. One alternative hypothesis might be that context, rather than familiarity, might modulate activity in the right hemisphere, and in particular in the right inferior frontal gyrus, during metaphor comprehension (cf. Diaz & Eppes, 2018). Still, it remains unclear why we found increased right inferior frontal gyrus activity during metonymy, but not during metaphor comprehension, since the context we provided was the same across conditions (i.e. the referent picture). One hypothesis might be that activity from the left inferior frontal gyrus spread toward its right homolog for resolving the representational conflict elicited by the metonymic allusions.

4.2.2. The frontoparietal control system downregulation during metaphor processing

Whereas metaphor, metonymy, and nonsensical sentences all upregulated the frontotemporal semantic network, other regions showed to be downregulated differently and depending on the experimental condition. The metaphorical and nonsensical conditions, when contrasted to the literal counterpart, elicited similar positive and negative activity patterns. We therefore hypothesized that the metaphorical sentences in our study might have been processed more similarly to nonsensical sentences. More in detail, brain areas belonging to the frontoparietal network (cingulate cortex, angular gyrus extending to the inferior parietal gyrus, prefrontal and frontal cortex) were found to be downregulated during metaphor processing. A similar pattern, although constrained to the cingulate cortex, was found for the nonsensical condition, but not during metonymy processing. Metaphoric and nonsensical target words had in common a significantly greater semantic distance from the literal counterpart than metonymies (*Table 2*). Hence we hypothesized this greater distance to drive the activity patterns we found. The regions that were downregulated have been shown to be involved in cognitive control and conflict resolution (e.g. Carter & van Veen, 2007; Seghier, 2012; Gold & Buckner, 2002; Milham et al., 2001). We therefore explained this negative brain activity of the frontoparietal control system during metaphor processing in terms of increased flexibility in searching nonliteral-literal mappings in the semantic space. In line with our hypothesis, reduced cognitive control has been found to favor problem solving and creativity (for a review, see Amer et al., 2016). Moreover, in a recent study, high creative individuals were found to be more accurate and faster when asked to decide whether two words (either literal, idiomatic, metaphoric, or unrelated) were related to each other (Kennett et al., 2018), suggesting that they were able to search the semantic space more functionally than low creative participants.

In the nonsensical condition, the downregulation included the left cingulate gyrus without, however, spreading toward parietal and frontal regions and across both hemispheres, like during metaphor processing. The lack of a spreader downregulation in the nonsensical condition could be explained by an overall lack of sensicality of the nonsensical target words in the context given by the sentence. This might have minimized the representational conflict as the conceptual domain of the target word fell too far apart from the context of the sentence. Accordingly, the activity maps for nonsensical > literal showed a frontotemporal upregulation that was inferior to the one elicited by the other contrasts of interest (*Figure 3*).

4.2.3. Theory-of-Mind and Language in metaphor and metonym comprehension

Metonymy vs metaphor processing elicited increased parietal activity in the bilateral precuneus and posterior cingulate regions. The cluster overlapped almost perfectly with part of the brain regions upregulated during the presentation of the False Beliefs stories in the ToM localizer task. In addition, the right inferior frontal gyrus was found to be active during the ToM task as well as during the metonymy processing. However, this region is not thought to play a key role in ToM processes (for a discussion, see Molenberghs, et al., 2016), hence we interpreted its upregulation in terms of world knowledge integration and representational conflict resolution, as previously argued.

The same regions that were found to be upregulated during metonymy processing and the ToM localizer task, were also downregulated during metaphor processing, suggesting the recruitment of additional ToM regions during metonymy processing.

The overlays between the activity blobs for metaphor/metonymy vs literal and each of the localizers, showed the involvement of regions belonging to the both networks (the left middle temporal gyrus, and right angular gyrus in particular). The degree of overlap between the language and the ToM localizer tasks, however, makes hard to disentangle the extent to which the same brain regions were upregulated by enhanced linguistic processing or inferential mechanisms. Looking at the individual activation maps for nonliteral language processing, language and ToM, might elucidate whether the same subject engaged in either of the two processes during metaphor and metonymy comprehension. Accordingly, it has been shown that there is a high inter-variability in pragmatic ability (Hagoort & Levinson, 2014). These individual differences could be mirrored at the neural level, thus making possible to disentangle the ToM from the language network.

5. Conclusions

In conclusion, the present study investigated the neural bases of metonymy and metaphor processing. In particular we were interested in whether the underlying neural mechanism of metonymy and metaphor comprehension differ and the extent to which the regions involved mirror inferential processes or increased language processing.

Our results suggest that similar neural mechanisms might be involved during metonymy and metaphor processing. Moreover, the activity patterns showed to be modulated by the extent of semantic distance between the target of the metaphoric/metonymic allusion and the metaphor/metonymy itself. The hypothesis is that metaphor processing might rely on a decreased semantic control for allowing the system to be more flexible in the semantic search space and thus be able to obviate the greater semantic distance between the source and the target domain. On the other hand, given the natural relatedness between metonymic allusions and their referent, metonymy resolution might rely only on the additional contribution of the right inferior frontal gyrus. This finding is important because it might suggest the involvement of cognitive control and conflict resolution during figurative language processing to a greater extent than previously thought.

In addition, the increased frontotemporal activity we found for both types of figurative language overlapped with regions belonging to the language network, as well as with some of the areas involved in inferencing and Theory-of-Mind. We conclude that this upregulation mirrored both inferential and enhanced linguistic processes. In future, it might be worthy to further investigate the role of Theory-of-Mind and language processing in metaphor and metonymy comprehension, for example by looking at subject-level contrast images across figurative conditions and at the extent of overlap with the individual activity patterns elicited by the ToM and language localizer tasks.

6. References

Ahrens, K, Liu, H, Lee, C., et al. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain and Language, 100(2),* 163-171.

Amer, T., Campbell, K. L., & Hasher, L. (2016). Cognitive control as a double-edged sword. *Trends in Cognitive Sciences, 20(12),* 905-915.

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Bach, K. (1994). Conversational Impliciture. *Mind & Language, 9(2),* 124-162.

Bambini, V., Bertini, C., Schaeken, W., et al. (2016). Disentangling metaphor from context: an ERP study. *Frontiers in Psychology, 7(559),* 1-14.

Boerger, M. A. (2005). Variations in figurative language use as a function of mode of communication. *Journal of Psycholinguistic Research, 34(1),* 31-49.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 5.3.55, retrieved from http://www.praat.org/

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brain behind figurative language: a quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia, 50(11),* 2669-2683.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience, 7(4),* 267-279.

De Grawe , S., Swain, A., Holcomb, P. J., et al. (2010). Electrophysiological insights into the processing of nominal metaphors. *Neuropsychologia, 48(7),* 1965-1984.

Diaz, M. T., & Eppes, A. (2018). Factors influencing right hemisphere engagement during metaphor comprehension. *Frontiers in Psychology, 9(414),* 1-6.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., et al. (2011). fMRI item analysis in a theory of mind task. *Neuroimage, 55*, 705-712.

Eizaga Rebollar, B. (2015). A relevance-theoretic perspective on metonymy. *Procedia – Social and Behavioral Sciences, 173,* 191-198.

Gold, B. T., & Buckner, R. L. (2012). Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron, 35,* 803-812.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4(416),* 1-13.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9(9),* 416-423.

Hagoort, P., Hald, L., Bastiaansen, M., et al. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304,* 438-441.

Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In: M. S. Gazzaniga, & G. R. Mangunpp (Ed.), *The Cognitive Neurosciences, 5th Edn* (pp. 667–674). Cambridge, MA: MIT Press.

Indefrey, P., & Cutler, A. (2004). Prelexical and lexical processing in listening. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences, 3rd Edn* (pp. 759-774). Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackson, R., Hoffman, P., Pobric, G., et al. (2016). The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience, 36(5),* 1490-1501.

Jang, G., Yoon, S., Lee, H., et al. (2013). Everyday conversations requires cognitive inference: Neural bases of comprehending implicated meanings in conversations. *Neuroimage, 81,* 61-72.

Kennett, Y. N., Levy, O., Kenett, D. Y., et al. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. *PNAS, 115(5),* 867-872.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42(3),* 643–650.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff , G., & Turner, M. (1989). *More than cool reason*. Chicago: University of Chicago Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.

Lam, N. H. L., Schoffelen, J., Uddén, J., et al. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *Neuroimage, 142,* 43-54.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92,* 57-78.

Mazaika, P. K., Hoeft, F., Glover, G. H., et al. (2009). Methods and Software for fMRI Analysis of Clinical Subjects. *Neuroimage 47,* S58.

Menenti L., Petersson K. M., Scheeringa R., et al. (2009). When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience, 21,* 2358–2368

Milham, M. P., Banich, M. T., Webb, A., et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain research, 12(3),* 467-473.

Mio, J. S., & Katz, A. N. (1996). *Metaphor: implications and applications*. Sussex: Psychology Press.

Molenberghs, P., Johnson, H, Henry, J. D., et al. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Behavioral Reviews, 65,* 276-291.

Novick, J. M., Kan, I. P., Trueswell, J. C., et al. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology, 26(6),* 527-567.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3(2),* 143-184.

Pinango, M. M., Zhang, M., Foster-Hanson, E., et al. (2016). Metonymy as referential dependency: Psycholinguistic and neurolinguistic arguments for a unified linguistic treatment. *Cognitive Science, 41,* 351-378.

Pynte, J., Besson, M., Robichon, F., et al. (1996). The time-course of metaphor comprehension: an event-related potential study. *Brain and Language, 316*, 293-316.

Rapp, A. M, Erb, M, Grodd, W., et al. (2011). Neural correlates of metonymy resolution, *brain & Language, 119,* 196-205.

Rapp, A. M., Mutschler, D. E., & Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage, 63(1),* 600-610.

Schumacher, P. B. (2013). When combinatorial processing results in reconceptualization: toward a new approach of compositionality. *Frontiers in Psychology, 4(677),* 1-13.

Schumacher, P. B. (2014). Content and context in incremental processing: "The ham sandwich" revisited. *Philosophical Studies, 168(1),* 151-165.

Seghier, M. L. (2012). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist, 19(1),* 43-61.

Song, S. (2011). Metaphor and Metonymy – A tentative Research into Modern Cognitive Linguistics. *Theory and Practice in Language Studies, 1(1),* 68-73.

Stolk, A, Noordzij, M. L., Verhagen, L., et al. (2014). Cerebral coherence between communicators marks the emergence of meaning. *PNAS, 111(51),* 18183-18188.

Thothathiri, M., Kim, A., Trueswell, A. K., et al. (2012). Parametric effects of syntactic-semantic conflict in Broca's area during sentence processing. *Brain & Language, 120,* 259-264.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavioral Research Methods, 38(4),* 584-589.

Wayne, D. (2014). Implicature. In E. Zalta (Ed.). *Stanford Encyclopedia of Philosophy*. Retrieved from https://plato.stanford.edu/entries/implicature/

Weiland, H., Bambini, V., & Schumacher, P. B. (2014). The role of literal meaning in figurative language comprehension: evidence from masked priming ERP. *Frontiers in Human Neuroscience, 8(583),* 1-17.

Yang, F.G., Fuller, J., Khodaparast, N., et al. (2010). Figurative language processing after traumatic brain injury in adults: a preliminary study. *Neuropsychologia, 48(7),*1923-1929.