Metaphorical language processing and amygdala activation in L1 and L2

Francesca M.M. Citron[1], Nora Michaelis[2], Adele E. Goldberg[3]

[1]Department of Psychology, Lancaster University, UK

[2]Freie Universität Berlin, Germany

[3]Psychology Department, Princeton University, NJ, USA

Address correspondence to:
Francesca Citron
Department of Psychology
Fylde College
Lancaster University
LA1 4YF
Lancaster, UK

Telephone: +44 1524594573
E-mail: fmm.citron@gmail.com

Running head: Metaphors and the amygdala in L2 processing

Abstract

The present study aims to investigate the neural correlates of processing conventional figurative language in non-native speakers in a comparison with native speakers. Italian proficient L2 learners of German and German native speakers read conventional metaphorical statements as well as literal paraphrases that were comparable on a range of psycholinguistic variables. Results confirm previous findings that native speakers show increased activity for metaphorical processing, and left amygdala activation increases with increasing Metaphoricity. At the whole-brain level, L2 learners showed the expected overall differences in activation when compared to native speakers (in the fronto-temporal network). But L2 speakers did not show any distinctive activation outside the caudate nucleus as Metaphoricity increased, suggesting that the L2 speakers were less affected by increasing Metaphoricity than native speakers were. With small volume correction, only a single peak in the amygdala reached threshold for L2 speakers as Metaphoricity increased. The findings are consistent with the view that metaphorical language is more engaging for native speakers but not necessarily for L2 speakers.

Keywords: multilingualism, metaphor, amygdala, emotion, fMRI, IFG

1 Introduction

Conventional metaphors are pervasive in everyday communication (Cameron, 2008; Pollio, Barlow, Fine, & Pollio, 1977), and they are sometimes so familiar they are barely recognized as metaphorical. Why is metaphorical language so commonly used, even when literal paraphrases exist? Potential aesthetic and rhetorical advantages of metaphorical language have intrigued rhetoricians as far back as Aristotle, yet until relatively recently, possible affective differences between metaphorical and literal language have been relatively unexplored by neuroscientists (Schmidt, Kranjec, Cardillo, & Chatterjee, 2010). Even less is known about such potential differences in the case of second language (L2) speakers' processing of metaphors. And yet half of the world's population uses an L2 in their daily lives (Bialystok, Craik, & Luk, 2012).

Metaphors have been argued to be a powerful persuasive tool in advertising (Chang & Yen, 2013; McQuarrie & Mick, 1996) and in political discourse (Charteris-Black, 2011; Semino, 2008). Even quite conventional metaphorical expressions have been found to increase persuasiveness and perceived creativity, while decreasing perceived complexity (Burgers, Konijn, Steen, & Iepsma, 2015). Conventional metaphorical expressions can give rise to inferences that do not necessarily hold of their literal counterparts. For example, if we describe someone as *over the hill* rather than *old,* we are more likely to infer that the person was once active and engaged, and that they are unlikely to continue to be so (Bowdle & Gentner, 2005; Carston, 2012; Gibbs, 2011; Thibodeau, Hendricks, & Boroditsky, 2017).

Metaphors appear to be useful in capturing emotions as they are preferred over literal expressions when people describe *how they felt* during an autobiographical event compared to *what happened* during the same event, and the preference is even stronger if the event is particularly emotionally intense (Fainsilber & Ortony, 1987; Ortony & Fainsilber, 1987; see also Crawford, 2009). Furthermore, the use of metaphors creates a stronger sense of perceived intimacy between the protagonists of a story, and enhances theory of mind (Bowes & Katz, 2015; Horton, 2007).

Certain differences between literal and metaphorical language processing have been reliably reported for native speakers. Reading conventional metaphors has been found to elicit significantly enhanced activation of the left inferior frontal gyrus (IFG), associated with inhibition, working memory, and executive functions more generally, and the left superior temporal gyrus (STG), which indexes semantic processing (Bohrn, Altmann, & Jacobs, 2012; Lai, van Dam, Conant, Binder, & Desai, 2015; Rapp, Mutschler, & Erb, 2012; Yang, 2014). The increased recruitment of the IFG and STG is evident even though reaction times often show no difference in speed between the processing of conventional metaphorical and literal stimuli (Gibbs, Lima, & Francozo, 2004; Giora, 1999; Glucksberg, 1998; Keysar, 1989) unless a sensicality judgment task is used in which case slower responses to conventional metaphors have been revealed (Bambini, Ghio, Moro, & Schumacher, 2013; Lai & Curran, 2013; Lai, Curran, & Menn, 2009)[1]. The increase in prefrontal and temporal regions is evident even when quite conventional metaphors are processed, and even when familiarity and complexity are taken into account (Citron & Goldberg, 2014; Citron et al., 2016b). Furthermore, EEG work shows time-course differences whereby conventional metaphors elicit larger N400 amplitudes than literal sentences, indexing contextual expectations (Bambini, Bertini, Schaeken, Stella, & Di Russo, 2016; Coulson & Van Petten, 2002; Weiland, Bambini, & Schumacher, 2014) and/or activation of metaphorical mappings (Lai & Curran, 2013; Lai et al., 2009), and larger P600 effects, indexing interpretation and pragmatic integration processes that are independent of context (Bambini et al., 2016; Coulson & Van Petten, 2002; Weiland et al., 2014); these differences are also not necessarily accompanied by reaction time differences (Weiland et al., 2014).

FMRI work on possibly distinctive *affective* responses to metaphorical processing is still in its infancy. The present work follows up on a recent finding of significantly enhanced activation of the left amygdala when native speakers silently read conventional metaphorical formulations, e.g., *She looked at him* <u>*sweetly*</u>, compared to almost identical literal paraphrases, i.e., *She looked at him*

---

[1] In particular, sensicality judgments were more costly in terms of accuracy and reaction times for metaphorical than literal statements (Bambini et al., 2013).

kindly (Citron & Goldberg, 2014). Metaphorical and literal formulations were explicitly rated as highly similar in meaning and as having the same degree of emotional content. Given that the amygdala response is functionally associated with the processing of evolutionary relevant or contextually salient (emotional) stimuli (Cunningham & Brosch, 2012; Garavan, Pendergrass, Ross, Stein, & Risinger, 2001; Hamann & Mao, 2002; Seeley et al., 2007), we interpreted the results as evidence of stronger emotional engagement elicited by metaphorical renderings.

The stimuli used in Citron and Goldberg (2014) all contained metaphors involving the source domain of *taste*, which may be more affectively loaded than other perceptual domains (Winter, 2016), but the increase in left amygdala activation for metaphorical language has since been replicated during the reading of naturalistic longer passages that included a range of different conventional metaphorical mappings (Citron et al., 2016b). This finding is additionally supported by a meta-analysis of 23 neuroimaging studies comparing figurative and literal materials (Bohrn et al., 2012), by an fMRI study comparing metaphorical and literal compound words (Forgács et al., 2012), and by converging evidence showing enhanced heart rate response (a measure of physiological arousal) for metaphorical translations of English metaphors into Spanish than literal translations (Rojo, Ramos, & Valenzuela, 2014).

As many people regularly rely on their L2 to communicate (Bialystok et al., 2012), it is important to gain a better understanding of potential emotional or rhetorical effects of figurative language in non-native speakers, given that metaphorical language is pervasive and impactful for L1 speakers. Metaphorical language competence is generally not considered a core skill in L2 learning and teaching (Littlemore & Low, 2006). Hence, even highly proficient L2 speakers struggle to understand and use metaphorical expressions effectively in their L2, find these difficult even if they contain familiar words, and are often unaware of their misinterpretations (Littlemore, Chen, Koester, & Barnden, 2011). In the specific case of idioms, proficient L2 speakers represent the literal (i.e., less salient) meaning, even when they know its idiomatic meaning, and even if the idiom is embedded in a figurative context (Cieślika, 2006; Cieślika & Heredia, 2011; Mashal, Faust, Hendler, & Jung-Beeman, 2008).[2] At the same time, L2 speakers benefit from similarity between metaphorical language in L1 and L2 by showing better comprehension of L2 metaphors which have corresponding metaphorical interpretations in their L2, particularly when no supportive context is provided (Türker, 2016).

Supporting the idea that L2 speakers often treat conventional metaphors as more novel than L1 speakers do is a study of metaphorical word pairs by Mashal, Borodkin, Maliniak and Faust (2015). While native speakers processed conventional metaphors faster when presented to the left (LH) than the right hemisphere (RH), L2 speakers showed the opposite effect: namely, conventional metaphors were processed as if they were less salient or “more novel”, i.e., faster when presented in the RH than the LH (Mashal et al., 2015). That is, the RH is recognized to play a key role in the processing of less salient figurative expressions (e.g., novel metaphors) as well as literal expressions (e.g., unusual literal word pairs or induced literal interpretation of ambiguous idioms) (Cardillo, Watson, Schmidt, Kranjec, & Chatterjee, 2012; Forgács, Lukács, & Pléh, 2014; Kasparian, 2013; Mashal & Faust, 2008; Mashal, Faust, & Hendler, 2005; Yang, 2014). In ERP work, late bilinguals have been found to respond more similarly to novel and conventional metaphorical word pairs in their L2, exhibiting similar amplitude of a late positivite component; a clear distinction between the two conditions was visible on the same component in their native language instead (Jankowiak, Rataj, & Naskręcki, 2017; see also Siyanova-Chanturia, Canal, & Heredia, 2019 for a review of ERP work on multilinguals).

In terms of affective responses, comprehension in L2 may be less emotionally engaging than in one’s native language. For example, taboo words are recognized by speakers to feel less impactful in L2 than in L1 (Dewaele, 2004); they are also less distracting in L2 than L1 (Colbeck & Bowers, 2012), and evoke less skin conductance, typically associated with emotion, in L2 than in

---

[2] Idioms and conventional metaphors are overlapping categories, as many conventional metaphors are also idioms (e.g., *swimming upstream, going with the flow, hold your horses)*. Yet single word metaphorical expressions are not generally considered idioms (e.g., *sweet* to mean “kind”), and there exist idioms that are not metaphorical in that they do not involve two distinct semantic domains *(e.g., by and large* or *the more, the merrier*).

L1 (Harris, 2004; Harris, Ayçiçeği, & Gleason, 2003). Other behavioral work has found participants form less clear mental imagery on the basis of cues provided in L2 (Hayakawa & Keysar, 2018) and are less affected by emotion in decision making as well (Hayakawa, Tannenbaum, Costa, Corey, & Keysar, 2017). A study by Hsu, Jacobs and Conrad (2015) on late, proficient L2 speakers of English (L1 German) showed equally good comprehension of happy, fearful and emotionally neutral Harry Potter passages in L1 and L2; however, reading in L1 elicited stronger activation of the emotion neural network including bilateral amygdala than reading in L2, and a better distinction between emotive and neutral passages was apparent in L1 than in L2 (Hsu et al., 2015).

Here we ask, do proficient L2 speakers, like native speakers, display an increase in amygdala activation for metaphorical language when compared to literal paraphrases? And more generally do L2 speakers process metaphorical language the same way that native speakers do?

For language processing in general, L2 comprehension involves the same network as L1 comprehension does, only more so (Briellmann et al., 2004; Chee, Tan, & Thiel, 1999; Perani & Abutalebi, 2005). That is, language comprehension requires a bilateral, fronto-temporal network including the IFG bilaterally, the temporal lobes, the dorso-medial prefrontal cortex (dmPFC) and the temporo-parietal junction (TPJ; Ferstl, 2010; Ferstl, Neumann, Bogler, & von Cramon, 2008; Mar, 2011). In L2 speakers, this network shows more widespread activations, and is often accompanied by the additional recruitment of pre-frontal regions (e.g., Briellmann et al., 2004). In addition, multilinguals recruit the 'switching network', which includes the supplementary motor area (SMA), the anterior cingulate cortex (ACC), associated with task and conflict monitoring, and the caudate nucleus, responsible for language selection (Abutalebi et al., 2013; Abutalebi & Green, 2007; Luk, Green, Abutalebi, & Grady, 2011).

The present fMRI study investigates how conventional metaphors are processed in both L1 and L2, with special attention to possible amygdala activation, in order to better understand possible affective effects of metaphorical language in both groups. The German metaphorical expressions used are compared with literal paraphrases that are matched on psycholinguistic and affective variables. A group of native German speakers was compared with a group of Italian speakers, living in Germany, who are highly proficient in German as a second language. We predict that, in accord with much previous work, during silent reading of both metaphorical and literal sentences, L2 speakers will show significantly enhanced activation of parts of the extended language network when compared with native speakers. Furthermore, in line with other results just reviewed, in response to metaphorical compared with literal sentences, we expected native speakers to show significantly enhanced activation of prefrontal regions including IFG, as well as STG, and left amygdala. The present study explores whether metaphorical language processing in L2 evokes the same increases in neural activity as it does in L1. Of particular interest is the relative activation of the left amygdala when metaphorical and literal sentences are compared in non-native speakers.

2 Method

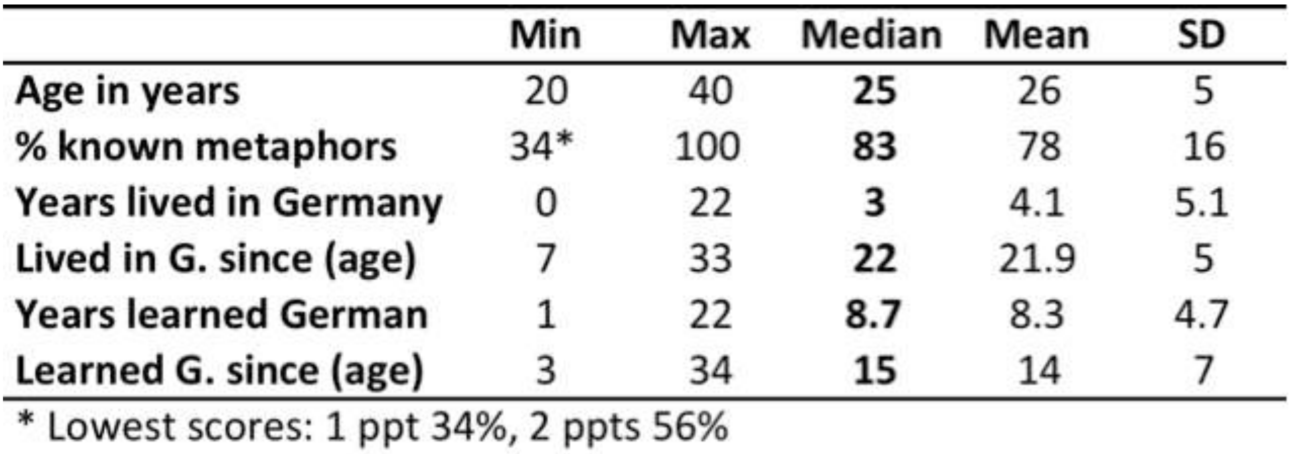
This study was approved by the Ethics Committee of the Freie Universität Berlin and is in accord with the guidelines of the American Psychological Association. The raw fMRI and behavioral data on L2 speakers are openly accessible at: https://openneuro.org/datasets/ds002221. The data on native speakers are openly accessible at: https://openneuro.org/datasets/ds002219. The data sharing adopted is in compliance with the requirements and the ethics approval of Lancaster University.

2.1 Participants

Twenty-five German native speakers from the Berlin area (15 women, 21-35 years, *mean age* = 26, *SD* = 4) and 22 proficient speakers of German whose native language is Italian (19 women, 20-40 years, *mean age* = 26, *SD* = 5) took part in the experiment. All Italian native speakers were living in Germany and were tested in German. They all considered themselves proficient speakers of German which was a requirement highlighted in the study's advertisement. In terms of language dominance, all participants had acquired Italian first then German, and only two participants learned German before 7 years of age. Therefore, except the latter two who may have similar dominance between the two languages, all had Italian as dominant language although they were immersed in a German-speaking country. After the experiment, L2 speakers typed the meaning of each metaphor previously presented in an online survey. Two independent judges determined whether their answers were correct or not. Most L2 speakers showed knowledge of the meaning of most metaphors: median accuracy 83%. We explored relationships between knowledge, years lived in Germany, age since which L2 speakers lived in Germany, how long they learned German for, and since what age (Table 1). Knowledge of the metaphors correlated significantly only with years lived in Germany, $r = 0.53$, $p < .01$. All remaining variables were highly correlated with one another; the only exception was the absence of correlation between years lived in Germany and the age at which L2 speakers started to learn German (Table 2).

All participants had normal or corrected-to-normal vision, and no neurological diseases or learning disabilities. Participants were each paid 20€. They all gave informed consent prior to the experiment.

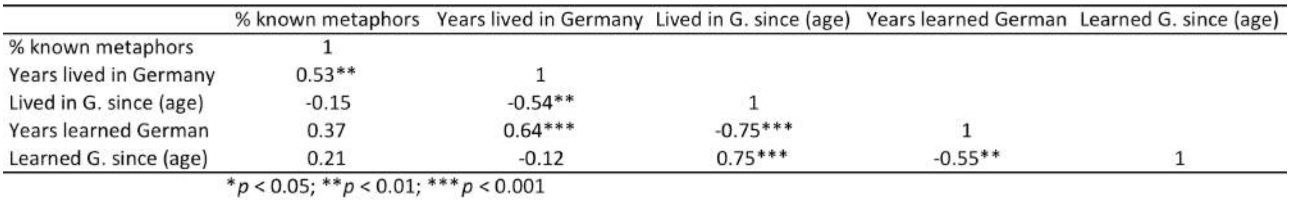
Table 1. Descriptive statistics of L2 speakers' age, knowledge of the meaning of the metaphors presented, and other variables related to their knowledge of and exposure to the German language.



	Min	Max	Median	Mean	SD
Age in years	20	40	**25**	26	5
% known metaphors	34*	100	**83**	78	16
Years lived in Germany	0	22	**3**	4.1	5.1
Lived in G. since (age)	7	33	**22**	21.9	5
Years learned German	1	22	**8.7**	8.3	4.7
Learned G. since (age)	3	34	**15**	14	7

* Lowest scores: 1 ppt 34%, 2 ppts 56%

Table 2. Pearson's correlations between L2 speakers' knowledge of the metaphorical meaning and their knowledge of and exposure to German.



	% known metaphors	Years lived in Germany	Lived in G. since (age)	Years learned German	Learned G. since (age)
% known metaphors	1				
Years lived in Germany	0.53**	1			
Lived in G. since (age)	-0.15	-0.54**	1		
Years learned German	0.37	0.64***	-0.75***	1	
Learned G. since (age)	0.21	-0.12	0.75***	-0.55**	1

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

None of our participants were monolingual speakers as the German native speakers all knew English (and the Italian speakers knew German). Because our focus was on a comparison of L1 and L2, we did not investigate possible effects of additional languages on language processing (but see Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Dijkstra & van Heuven, 2002; Lehtonen et al., 2012).

2.2 Materials

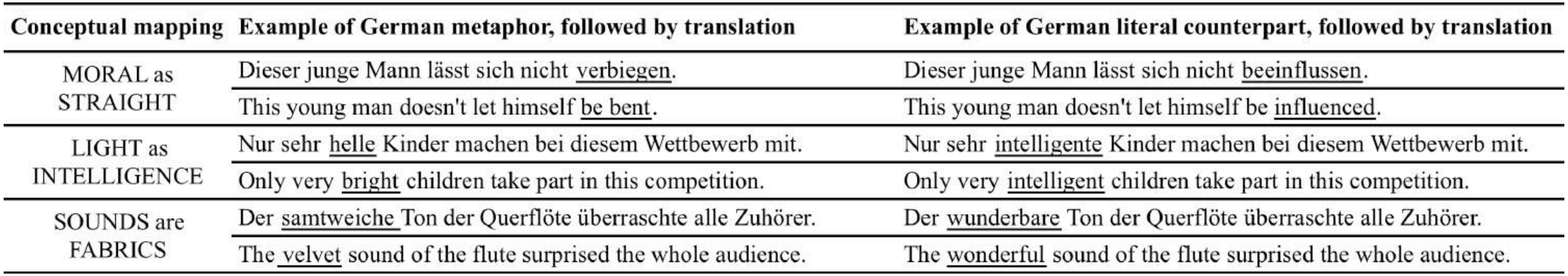
Thirty-two German conventional metaphorical sentences (MSE) were selected from the COMETA dataset (Citron, Lee, & Michaelis, 2019b) and included a single word (e.g., "light") that would be interpreted metaphorically, given the sentential context; 32 literal sentences (LSE) would contain its literal counterpart (e.g., "easy"; *Ach, was für ein <leichtes> / <einfaches> Leben hatten wir während der Schulferien!* "Oh, what a <light> / <easy> life we had during the school holiday!"). Each metaphorical sentence relied on a different conceptual mapping (WEIGHT as COMPLEXITY in the above example); hence, no unique concrete domain was targeted across trials (see examples in Table 3). MSE and LSE were rated as equally imageable, familiar[3], emotionally valenced and arousing (all *ts* (31) < 1.81, *ns*), and as highly similar in meaning (see Table 4 for descriptive statistics); they were also matched for length in letters and words (*ts* (62) < 0.75, *ns*). As intended MSE were rated as significantly higher in metaphoricity than LSE ($t(31) = 9.54, p < .001$). Thirty-two hash marks strings (HMS) were created as a baseline, similar to the sentences in length and number of continuous sequences (e.g., ## ###### #### ###### ###). MSE, LSE and HMS were matched for length in words/sequences and letters/hash marks. Ten yes/no comprehension questions were created. In addition, 4 filler sentences were included.

Certain metaphorical stimuli had closer paraphrases in Italian. In order to include this factor in our analysis we used the following measure of similarity : 1 = same conventional expression in L1: 12 expressions, 37.5%; 2 = similar conventional metaphor in L1: 8 expressions, 25%; 3 = one may say it but it's not a conventional expression: 4 expressions, 12.5%; 4 = does not exist in L1: 6 expressions, 18.8%; 5 = does not exist in L1 and is very difficult to imagine: 2 expressions, 6.3%. Hence, most expressions were similar, with a mean rating of 2.31, *median* = 2, *SEM* = 0.24.

---

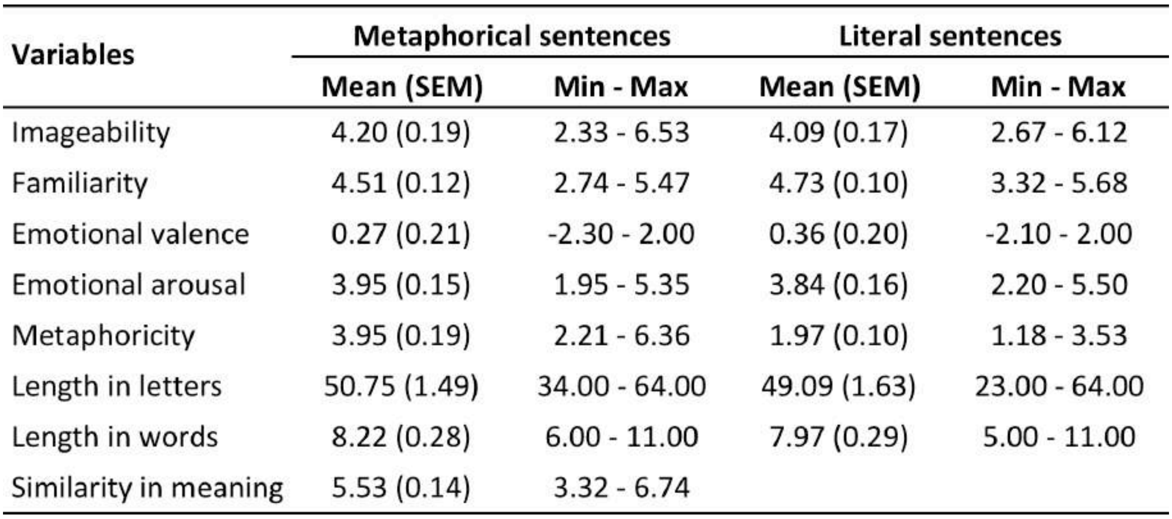
[3] We used familiarity (i.e., subjective frequency) ratings of the whole sentence for our stimuli instead of calculating the mean frequency of each word in the sentences because it has been demonstrated that familiarity ratings (or subjective frequency) are a better predictor of idiom processing (Bonin, Méot, & Bugaiska, 2013; Libben & Titone, 2008). Moreover, we were interested in familiarity of the overall expressions and some of the words are used less frequently in their literal senses (e.g., the German word, *samtweich,* "velvet" in Table 3).

Table 3. Examples of original German sentences used, followed by an English translation. Some of the metaphors employed do not exist in English, however we translated them literally so the reader can get an idea of the way in which the underlying conceptual mapping is realized. The metaphorical words and their literal counterparts are underlined.



Conceptual mapping	Example of German metaphor, followed by translation	Example of German literal counterpart, followed by translation
MORAL as STRAIGHT	Dieser junge Mann lässt sich nicht verbiegen.	Dieser junge Mann lässt sich nicht beeinflussen.
	This young man doesn't let himself be bent.	This young man doesn't let himself be influenced.
LIGHT as INTELLIGENCE	Nur sehr helle Kinder machen bei diesem Wettbewerb mit.	Nur sehr intelligente Kinder machen bei diesem Wettbewerb mit.
	Only very bright children take part in this competition.	Only very intelligent children take part in this competition.
SOUNDS are FABRICS	Der samtweiche Ton der Querflöte überraschte alle Zuhörer.	Der wunderbare Ton der Querflöte überraschte alle Zuhörer.
	The velvet sound of the flute surprised the whole audience.	The wonderful sound of the flute surprised the whole audience.

Table 4. Descriptive statistics of psycholinguistic and affective variables of sentences. Imageability, familiarity, emotional arousal, metaphoricity, and similarity in meaning range from 1 (not at all imageable, familiar, etc.) to 7 (very much). Emotional valence ranges from -3 (very negative), through 0 (neutral) to +3 (very positive).



Variables	Metaphorical sentences		Literal sentences	
	Mean (SEM)	Min - Max	Mean (SEM)	Min - Max
Imageability	4.20 (0.19)	2.33 - 6.53	4.09 (0.17)	2.67 - 6.12
Familiarity	4.51 (0.12)	2.74 - 5.47	4.73 (0.10)	3.32 - 5.68
Emotional valence	0.27 (0.21)	-2.30 - 2.00	0.36 (0.20)	-2.10 - 2.00
Emotional arousal	3.95 (0.15)	1.95 - 5.35	3.84 (0.16)	2.20 - 5.50
Metaphoricity	3.95 (0.19)	2.21 - 6.36	1.97 (0.10)	1.18 - 3.53
Length in letters	50.75 (1.49)	34.00 - 64.00	49.09 (1.63)	23.00 - 64.00
Length in words	8.22 (0.28)	6.00 - 11.00	7.97 (0.29)	5.00 - 11.00
Similarity in meaning	5.53 (0.14)	3.32 - 6.74		

2.3 Procedure

The experiment was conducted at the Center for Cognitive Neuroscience Berlin (C.C.N.B.), at the Freie Universität Berlin, and programmed with Presentation (Neurobehavioral System Inc.). Stimulus order and timings were optimized to maximize the statistical efficiency of the task design by using OPTSEQ2 (Dale, 1999), which created randomized sequences of experimental conditions and null events of varying durations (i.e., jittered). The stimuli were presented in 2 different runs: each run contained 2 filler sentences at the beginning, followed by 16 MSE, 16 LSE, 16 HMS, and 5 questions, in randomized order.

Participants read written instructions describing the whole experiment, signed the informed consent form and were led into the scanner room. First, the magnitude and phase images of the magnetic field in the scanner were measured (1 minute). Then, the experimenter repeated the task instructions orally, asking participants to silently read sentences for comprehension, to attend to the HMS, and to respond to occasional yes/no questions by pressing one of two buttons with their right index and middle fingers. Each functional scanning (or run) lasted 7 minutes (215 functional volumes acquired). After the reading task, a structural image was acquired (5 minutes).

Each stimulus was presented centrally on a computer monitor and projected into fMRI-compatible glasses, in white font on a black background. All sentences and HMS were presented for 4 seconds, whereas questions for 6 seconds; jittered inter-stimulus intervals (ISIs) varied between 1 and 7 seconds, during which time a fixation cross was centrally presented to keep participants' gaze and attention focused. The experiment lasted approximately 1 hour, including preparation, scanning and debriefing.

After the experiment, L2-speakers were asked about their formal knowledge of and exposure to German by living in a German-speaking country. They also typed the meaning of each metaphor they read during the experiment in an online questionnaire. Two independent judges determined whether these corresponded to the actual meaning of each metaphor. Importantly, based on this, all unknown or wrongly defined metaphors were excluded from the analyses of brain activity.

2.4 MRI data acquisition and pre-processing

Magnetic resonance images were acquired by means of a 3-Tesla Tim-Trio scanner (Siemens, Erlangen) equipped with a 12-channel receive RF head coil. Magnitude and phase images (field map) were first acquired: 37 slices per image; 3-mm thick with a 60˚ flip angle; voxel size: 3x3x3 mm; FOV 192 mm isotropic voxels without gap; matrix per slice: 64×64 mm; TR 488 ms; 2 TE: 4.92; 7.38 ms; acquisition time 1'05". For functional images, a standard EPI sequence was used, with following parameters: 37 slices, 3-mm thick with a 70˚ flip angle; voxel size: 3x3x3 mm; FOV 192 mm isotropic voxels without gap; matrix per slice: 64x64 mm; TR 2000 ms; TE 30 ms; acquisition time 8'36". Finally, full-brain, T1-weighted structural scans were acquired (MPRAGE sequence): 176 slices, 9˚ flip angle, voxel size: 1x1x1 mm, FOV 256 mm without gap; matrix per slice: 256×256 mm; TR 1900 ms, TE 2.52 ms, acquisition time 4'26". During acquisition of structural and functional images, six head movement parameters were continuously recorded: linear movements or translation, along the x y and z axes, and non-linear movements or rotation, including pitch, roll and yaw.

Processing of functional images and statistical analyses were performed using SPM12 (Welcome Trust Centre, http://www.fil.ion.ucl.ac.uk/spm), employing slice timing correction, realignment (images were re-aligned according to the 6 head-movement parameters recorded), unwarping (through the creation of a field map), and sequential co-registration to structural T1 images. Structural images were segmented into grey matter, white matter, cerebrospinal fluid, bone, soft tissue and air/background. Based on the segmented grey and white matter images, a group anatomical template was created with the DARTEL toolbox (Ashburner, 2007). Based on these transformation parameters, the functional images were then iteratively normalized to standard space (Montreal Neurologic Institute, MNI). Subsequently, functional volumes were spatially smoothed with a 6-mm Gaussian kernel to adjust for between-participants anatomical differences. Because no participant showed head movements larger than 3 mm, no additional procedures for head movement correction nor the exclusion of any participant from the analyses were necessary.

2.5 Statistical analyses

A General Linear Model was used in an event-related design. Hemodynamic responses were time-locked to the stimulus onset for the whole duration of each stimulus presentation and convolved with the canonical hemodynamic response function of SPM12.

2.5.1 Metaphorical vs. literal sentences. In this analysis, two conditions were compared in a factorial design: metaphorical vs. literal sentences. At the first level, for each participant 5 separate regressors were used to model each condition: MSE, LSE, HMS, questions, fillers. Finally, 6 head movement regressors were included in each model. At the second level, a mixed, 2x2 ANOVA with between-participants factor Group (L1 speakers, L2 speakers) and within-participants factor Sentence (MSE, LSE) was conducted. Contrasts for main effects and interactions were defined in both directions: L2 > L1, L1 > L2, MSE > LSE, LSE > MSE; MSE in L1 and LSE in L2 > LSE in L1 and MSE in L2, LSE in L1 and MSE in L2 > MSE in L1 and LSE in L2.

2.5.2 Metaphoricity as continuous variable. In this analysis, no categorical distinction between metaphorical and literal sentences was made, but rather numeric values of metaphoricity (from not at all metaphorical, i.e., literal, to highly metaphorical) were used by creating one continuous variable, in a parametric design. At the first level, the first regressor, Sentence, coded all MSE and LSE indistinctly, while a second continuous regressor, Metaphoricity, contained the corresponding ratings for each single sentence; further regressors were HMS, questions, fillers, and the 6 movement regressors. The contrast increase in Metaphoricity was defined. At the second level, contrast images for increase in Metaphoricity were compared between L1 and L2, therefore making it possible to explore interactive effects of Group and Metaphoricity; contrasts in both directions were defined: L1 > L2, L2 > L1. If interactive effects were significant, further contrasts within L1 and L2 were defined. In order to investigate the main effect of Metaphoricity, a conjunction analysis was performed, combining increase in Metaphoricity for both L1 and L2 groups.

For significance levels, a standard voxel-level threshold of $p < .005$ uncorrected was chosen, along with a cluster-level threshold, corrected for false discovery rate (FDR), of $p < .05$ (Lieberman & Cunningham, 2009). In addition, a priori small-volume correction (SVC) on the amygdala, bilaterally, was applied to any contrast comparing MSE > LSE and to any contrast using Metaphoricity, if this region had not already been found significantly active at the whole-brain level. The SVC was based on Brodmann’s areas, as implemented in the WFU PickAtlas toolbox (Maldjian, Laurienti, Kraf, & Burdette, 2003). For the SVC analyses, a voxel-level threshold of $p < .001$ uncorrected was chosen, and family-wise error (FWE) correction was applied at the voxel level with a threshold of $p < .05$ (Bennett, Wolford, & Miller, 2009).

3 Results

3.1 Differences between L1 and L2

In the ANOVA (Group by Sentence), we found a significant main effect of group, whereby L2 speakers showed significantly enhanced bilateral activation of the SMA and of peri-central cortices including the left and right pre-central gyri, the left middle temporal gyrus (MTG), the right hippocampus and parahippocampal gyrus, compared to native speakers (Appendix A, Figure 1a). Activation of the basal ganglia, including the caudate nuclei and putamen was also found (Figure 1a). Finally, involvement of the occipital lobe and the cerebellum, bilaterally, was found.

In contrast, native speakers showed significantly enhanced activation of lateral frontal cortices including the right middle frontal gyrus (MFG), wide-spread activation of left temporal cortices including the superior temporal gyrus (STG) bilaterally, extending to the left posterior insula, and a cluster in the posterior portion of the right MTG, compared to L2 speakers (Appendix A). They also showed enhanced activation of occipital medial cortices, including cunei, pre-cunei, posterior cingulate cortex (PCC) and retro-splenial cortex (RSC; Figure 1b, Appendix A).

No significant main effect of Sentence and no interaction between Group and Sentence were found.

Figure 1a.

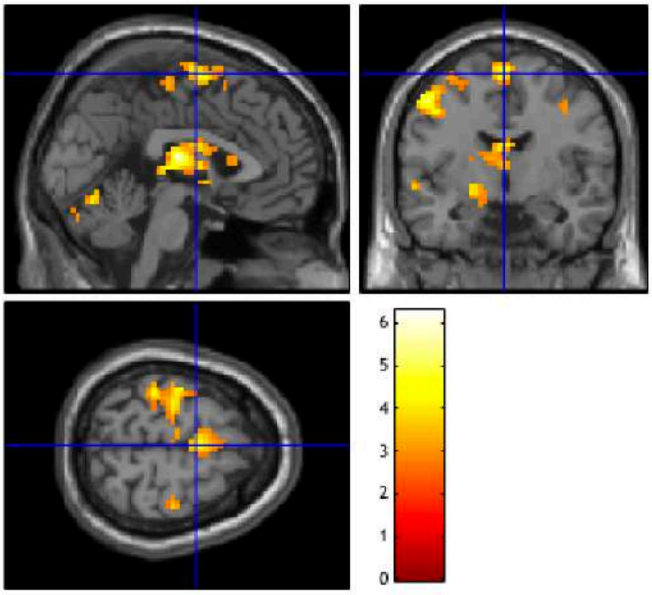


Figure 1b.

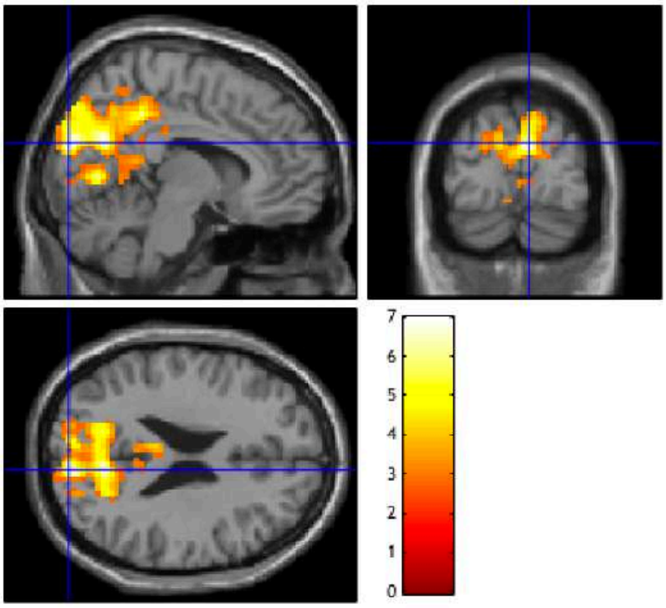


Figure 1a. Clusters of significantly enhanced activation in response to all sentences in L2 speakers compared to native speakers. The two clusters of interest include the SMA (MNI 0 -6 66) and the pre-central gyri, with the latter visible in the top right and bottom quadrants; however, other clusters of activation including the caudate nuclei and putamen are also visible at the bottom of the brain in the left top quadrant. **Figure 1b.** The opposite contrast shows clusters of significantly enhanced activation in posterior medial regions for native compared to L2 speakers, including the cunei (MNI -9 -87 24), pre-cunei, PCC and RSC.

3.2 Increase in Metaphoricity for native versus L2 speakers

Using a more fine-grained measure of metaphorical content in our sentences, i.e., increasing Metaphoricity ratings across metaphorical and literal stimuli, we found a significant interaction of this variable with Group. Specifically, when comparing L1 > L2 speakers by increasing Metaphoricity, we found significantly enhanced activation of medial frontal areas such as dmPFC and ACC, bilaterally, as well as the left dorsal SFG and right M/SFG (Table 5). No clusters of significant activation were found when comparing L2 > L1 by increasing Metaphoricity.

The significant interaction allowed us to explore increase in Metaphoricity within each group. At the whole-brain level, native speakers showed a pattern of increasing activations in the IFG bilaterally, extending to the right anterior insular cortex (AIC), the left amygdala (Table 5, Figure 2a), and medial frontal areas including the rostral ACC, ACC, and middle cingulate cortex (MCC; Figure 2b). Other frontal regions included right SMA and left MFG extending to dmPFC.

L2 speakers, on the other hand, only showed a significant increase in activation of the head of the right caudate nucleus correlating with the increase in Metaphoricity at the whole-brain level (Table 5, Figure 2c). A SVC on the amygdalae revealed only one voxel of significant activation in the left amygdala ($T = 3.53$, MNI = -21 -6 -15, part of a 3-voxel cluster).

Increasing Metaphoricity across groups revealed no significant clusters of activation. This result confirms the lack of a main effect of Sentence in the ANOVA, and suggests that L2 speakers may not distinguish metaphorical from literal materials or respond to an increasingly metaphorical sentential formulation the same way that L1 speakers do.

Because we had collected information about several individual-difference variables that could be expected to correlate with proficiency (Table 2), we conducted analyses of BOLD response in which each variable was used as a parametric regressor, to see whether any significant patterns of activation would appear, but no effects were evident, quite possibly because there was not enough variation in proficiency among our 22 participants. Furthermore, no significant patterns of brain activation were found in the L2 group in response to increasing or decreasing the similarity of the metaphorical expressions in German and Italian. In response to a reviewer's suggestion, we also re-ran our analyses excluding the three participants who misinterpreted the largest number of metaphors. The key results remain the same: As in the full data set, with increasing Metaphoricity, a significant cluster of activation in the right caudate nucleus was visible but no amygdala activity. With the small volume correction on the amygdala, we find a 3-voxel cluster of activation in the left amygdala, just as in the full data set, and an additional 2-voxel cluster in the right amygdala. This tiny new cluster is consistent with the idea that especially high proficiency may result in greater alignment between metaphorical processing in L2 and L1, a point we return to, although this weak evidence in the current dataset precludes further speculation. Finally, the main effect of group revealed very similar patterns of activation in the subset of data for both L1 > L2 and L2 > L1 contrasts, although most clusters were smaller in size.

Figure 2a.

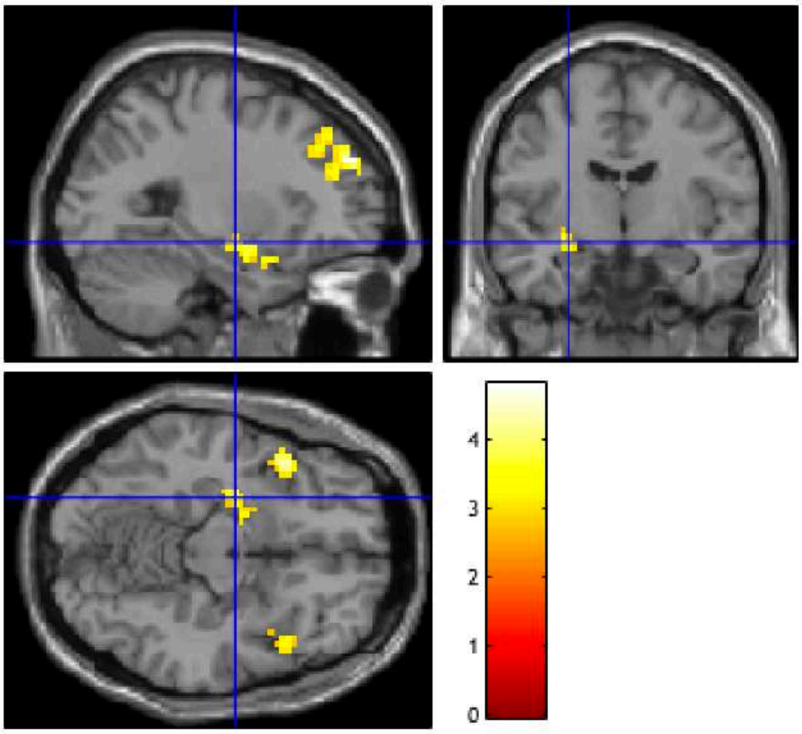


Figure 2b.

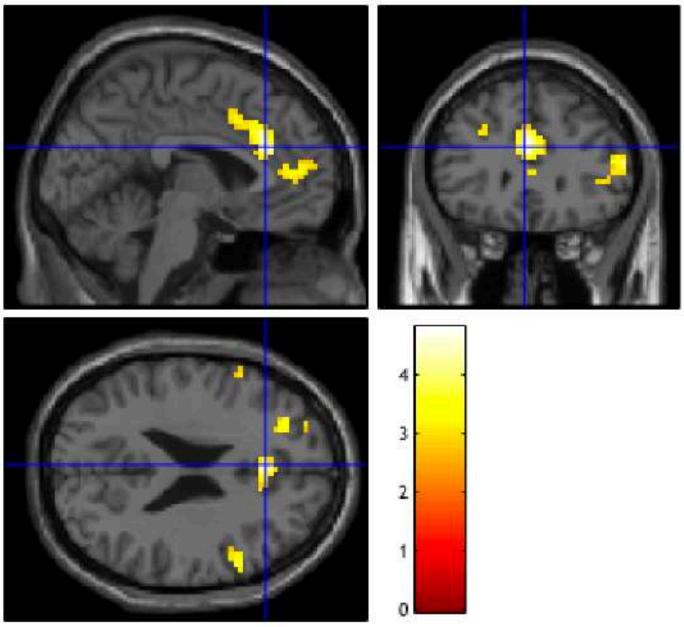


Figure 2c.

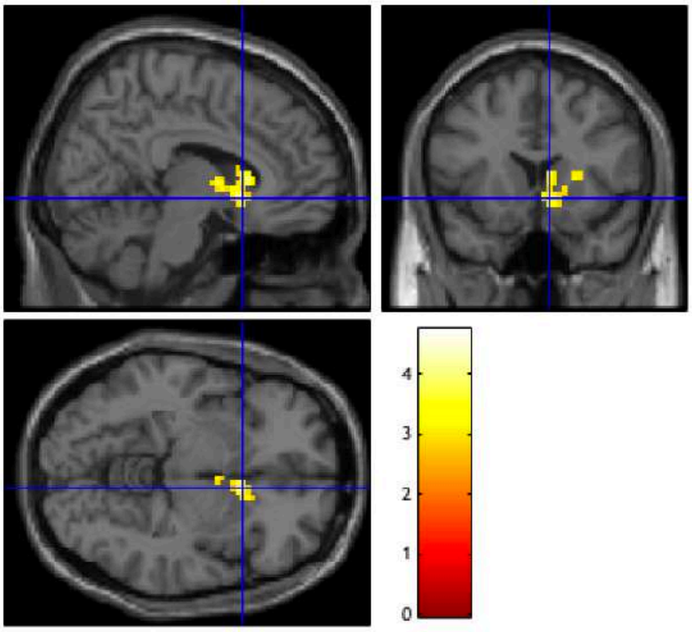
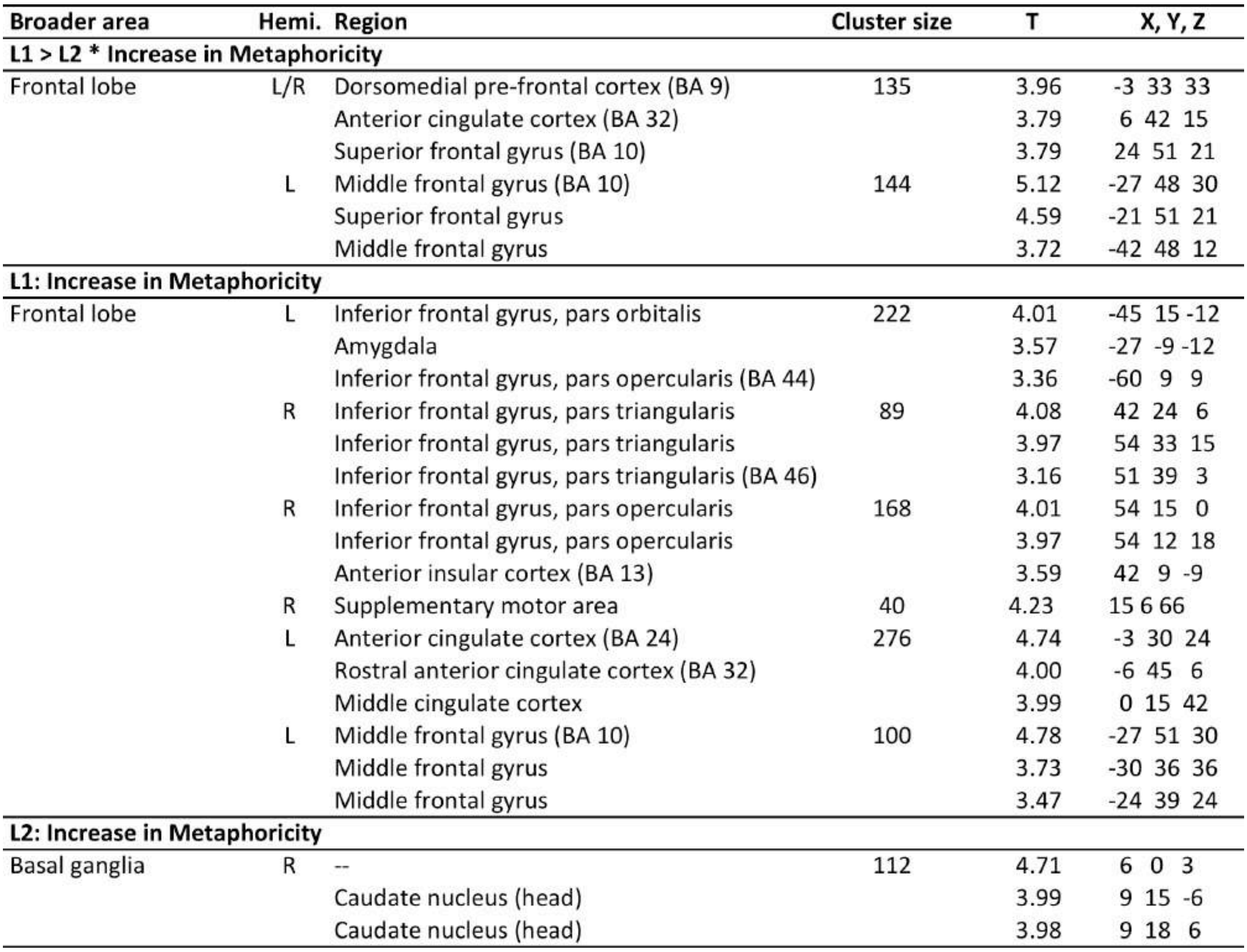


Figure 2. Clusters of significant activation for increase in Metaphoricity, within each group, at the whole-brain level. Native speakers show increasing activation in **(a)** the left amygdala (MNI -27 -9 -12); bilateral activation of the IFG is also visible in the bottom quadrant, and the left dmPFC and MFG are visible in the top left quadrant; **(b)** the ACC (MNI -3 30 24), extending to more rostral portions; **(c)** L2 speakers show instead only one cluster of activation in the right head of the right caudate nucleus (MNI 9 15 -6) at the whole-brain level.

Table 5. Clusters showing significant BOLD signal change at the whole-brain level for the interaction between increase in Metaphoricity and Group as well as increase in Metaphoricity within each group: native speakers (L1) and second language speakers (L2). At the voxel level, a significance threshold of $p < 0.005$ was applied, followed by FDR correction at the cluster level ($p < 0.05$). Legend: Hemi. = hemisphere, L = left, R = right; cluster size is in voxels, T = peak t value; X, Y, Z = MNI stereotactic space coordinates.



Broader area	**Hemi.**	**Region**	**Cluster size**	**T**	**X, Y, Z**
L1 > L2 * Increase in Metaphoricity					
Frontal lobe	L/R	Dorsomedial pre-frontal cortex (BA 9)	135	3.96	-3 33 33
		Anterior cingulate cortex (BA 32)		3.79	6 42 15
		Superior frontal gyrus (BA 10)		3.79	24 51 21
	L	Middle frontal gyrus (BA 10)	144	5.12	-27 48 30
		Superior frontal gyrus		4.59	-21 51 21
		Middle frontal gyrus		3.72	-42 48 12
L1: Increase in Metaphoricity					
Frontal lobe	L	Inferior frontal gyrus, pars orbitalis	222	4.01	-45 15 -12
		Amygdala		3.57	-27 -9 -12
		Inferior frontal gyrus, pars opercularis (BA 44)		3.36	-60 9 9
	R	Inferior frontal gyrus, pars triangularis	89	4.08	42 24 6
		Inferior frontal gyrus, pars triangularis		3.97	54 33 15
		Inferior frontal gyrus, pars triangularis (BA 46)		3.16	51 39 3
	R	Inferior frontal gyrus, pars opercularis	168	4.01	54 15 0
		Inferior frontal gyrus, pars opercularis		3.97	54 12 18
		Anterior insular cortex (BA 13)		3.59	42 9 -9
	R	Supplementary motor area	40	4.23	15 6 66
	L	Anterior cingulate cortex (BA 24)	276	4.74	-3 30 24
		Rostral anterior cingulate cortex (BA 32)		4.00	-6 45 6
		Middle cingulate cortex		3.99	0 15 42
	L	Middle frontal gyrus (BA 10)	100	4.78	-27 51 30
		Middle frontal gyrus		3.73	-30 36 36
		Middle frontal gyrus		3.47	-24 39 24
L2: Increase in Metaphoricity					
Basal ganglia	R	--	112	4.71	6 0 3
		Caudate nucleus (head)		3.99	9 15 -6
		Caudate nucleus (head)		3.98	9 18 6

4 Discussion

This study aimed to explore the neural correlates of conventional metaphor comprehension in speakers of a second language, by comparing them to native speakers, with particular attention to activation in the amygdala. Before focusing on the comparison of metaphorical and literal language, however, we first review general differences found between native and L2 speakers.

4.1 Language processing in L1 and L2

As expected, the task was more demanding for L2 speakers as they have had less experience with either sentence type, and greater activity was found in several areas during the silent reading of all literal and metaphorical sentences. L2 speakers showed significantly enhanced activation of the left middle and superior temporal gyri, associated with increasing semantic processing (Newmann, 2001). They also showed activation of regions associated with motor planning and coordination of speech, including the pre-central gyri and the basal ganglia (Dronkers, 1996; Price, 2012); these areas are often active during silent reading or written language comprehension in native speakers (Hagoort et al., 1999) and have been shown to be more strongly active during silent reading in L2 speakers than natives, likely due to their greater difficulty in motor planning (Rüschemeyer, Zysset, & Friederici, 2006). L2 speakers also showed stronger activation in bilateral visual areas, likely due to the greater demands required for reading in L2 (Koda, 1996), and in the right hippocampus and parahippocampal gyrus, suggesting enhanced encoding processes (Martin, 1999).

There was also some evidence of interference from L1 during L2 comprehension insofar as the executive control network was more active in the L2 group. In fact, the L2 group recruited the ‘switching network’, including SMA and ACC, associated with task and conflict monitoring, and the caudate nucleus, responsible for language selection (Abutalebi et al., 2013; Abutalebi & Green, 2007; Luk et al., 2011). This network is typically activated during (production) tasks that involve switching between different languages, and is more broadly related to conflict monitoring, maintenance of different representations in working memory, and ambiguity resolution (e.g., McNab et al., 2008; Osaka et al., 2004). We know from behavioral, eye-tracking, and psychophysiological studies that multilinguals activate both or all of their languages simultaneously to some extent, even during tasks that entail no conflict and involve only one language (Bialystok et al., 2012; Kroll, Bobb, Misra, & Guo, 2008; Marian, Spivey, & Hirsch, 2003; Thierry & Wu, 2007). Therefore, the switching network may be active even during silent reading in one language, and this is especially likely in L2 (but see Parker Jones et al., 2012 for contrasting evidence). This finding is consistent with other work showing interference from L1 during figurative language comprehension in L2 (Cieślika, 2006; Laufer, 2000; Türker, 2016).

As expected, native speakers displayed activation patterns that indicated easier processing when compared with L2 speakers. In particular, native speakers engaged the default mode network, typically active in rest conditions, associated with mind wandering and mentalizing (Buckner, Andrews-Hanna, & Schacter, 2008). This includes medial parieto-occipital areas such as the PCC, RSC, pre-cunei, and cunei, lateral parietal areas bilaterally such as the superior parietal lobules, angular gyri and SMG, and dorsal frontal areas such as the MFG and SFG (Andrews-Hanna, 2012; Buckner et al., 2008).

At the same time, native speakers appeared to engage in deeper semantic processing, possibly because the relatively light cognitive demands in the task allowed them more resources to retrieve and integrate meanings. The increase in semantic processing is evidenced by significantly enhanced extensive activation of the middle and superior temporal cortices bilaterally, extending to the right AIC; these regions represent the seat of our lexico-semantic and conceptual representations (Bookheimer, 2002; Mummery et al., 2000), are associated with semantic violations in sentences (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003), and integration/interpretation processes during text comprehension (Ferstl, 2010; Ferstl et al., 2008).

4.2 Metaphor processing in L1

Consistent with previous work, native speakers showed increased activation of several areas for metaphor comprehension when compared to literal language comprehension (Bohrn et al., 2012; Citron et al., 2016b; Rapp et al., 2012; Uchiyama et al., 2012; Yang, 2014). In particular, increasing Metaphoricity evoked increased activation in: IFG bilaterally, an index of executive control and maintenance of multiple representations in working memory (McNab et al., 2008; Osaka et al., 2004), and ACC, related to conflict monitoring, ambiguity resolution and emotional conflict (Botvinick, Nystrom, Fissell, Cater, & Cohen, 1999; Kanske & Kotz, 2011). Collectively, this activation pattern in prefrontal areas is suggestive of simultaneous activation of different lexico-semantic representations, i.e., the metaphorical and part of the literal meaning, along with selection of the intended meaning. Indeed, a link between individual differences in executive control and metaphor comprehension has also been reported in an eye-tracking study (Columbus et al., 2015).

Of particular interest is that native speakers displayed significant activation of the left amygdala at the whole brain level in response to increasing Metaphoricity ratings, a finding consistent with previous results on conventional taste metaphors and on naturalistic metaphorical stories (Citron & Goldberg, 2014; Citron et al., 2016b). This result serves to extend the link between metaphorical language and amygdala activation to simple metaphorical sentences that involve little explicit emotional content (and which are unrelated to taste).

4.3 Metaphor processing in L2

Differentiation between degrees of Metaphoricity in L2 speakers was not in evidence, except for one significant cluster in the head of the right caudate nucleus in response to increasing Metaphoricity rating; the caudate nucleus is part of the switching network and has been specifically associated with language selection. This finding is reminiscent of the interference from L1 reported during comprehension of L2 metaphors (Türker, 2016). Interestingly, there exists one neuroimaging study showing enhanced activation of the heads of the caudate nuclei for metaphorical short stories over their literal counterparts in native speakers; this was interpreted as indexing the restriction of sentence meanings between a few possible candidates (Uchiyama et al., 2012). It is possible this may not have been present in our native speakers due to the very high conventionality and simplicity of our stimuli. At the whole brain level, no other significant clusters of activation correlating with increased Metaphoricity in L2 were apparent. This, together with the significant interaction between group and Metaphoricity, indicates a difference between L2 and native speakers' processing of metaphorical language. In fact, when contrasting L1 > L2 with increasing Metaphoricity, significant activation of areas associated with conflict/ambiguity resolution (ACC), self-referential thought (dmPFC), and mentalizing (SMG, dorsal MFG) were found, while no significant clusters appeared in the opposite contrast: L2 > L1 for increasing Metaphoricity. The lack of increasing RH activation in the IFG, AIC or the temporal cortices may indicate that our L2 speakers treated the conventional metaphorical stimuli as conventional rather than novel, as intended (Cardillo et al., 2012; Forgács et al., 2014; Kasparian, 2013; Mashal & Faust, 2008; Mashal et al., 2005; Yang, 2014).

To summarize, at the whole-brain level native speakers showed greater activation in IFG, ACC, and left amygdala as Metaphoricity increased but no such activations were evident in L2 speakers. This suggests that proficient L2 speakers process conventional metaphors and literal sentences more similarly than L1 speakers do. Since we had a pre-defined interest in amygdala activation in particular, a small volume correction was performed within the group of L2 speakers. Only one voxel of significant activation was found (as part of a 3-voxel cluster), which in our study represents approximately 2.2% (27 mm^3) of amygdala volume (on average 1240 mm3, according to Brabec et al., 2010), which is relatively weak evidence for stronger emotional engagement with increasing Metaphoricity in L2 speakers. Thus, it appears that L2 speakers are equally engaged (or not) in response to literal and metaphorical sentences. That is, just as taboo words have been found to have less emotional impact in L2 compared to L1 (Harris, Gleason, & Ayçiçeği, 2006; Pavlenko, 2012), and reading fiction in L2 has been found to activate the neural network associated with emotion to a lesser extent than in L1 (Hsu et al., 2015), results here indicate that conventional metaphors are no more emotionally engaging in L2 than literal language, although they are for native speakers.

Two aspects of the current study deserve further attention in future work. First, while our L2 speakers considered themselves proficient in their L2 (German), it remains possible that non-native speakers at even higher proficiency levels will be found to process metaphorical language more like native speakers. This is in fact to be expected, given that much work has found that degree of proficiency, rather than age of acquisition, predicts language processing in behavioral & neural tasks (e.g., Ardal et al., 1990; Birdsong, 2005; Fabbro, 2001; Marinova-Todd, Marshall, & Snow, 2000; Perani et al., 1998). Second, our interest was on conventional metaphors because they pervade speakers' everyday language in ways that are hard to avoid (Lakoff & Johnson 1980). Because of this, we took care to create metaphorical stimuli that were recognized by native speakers and interpreted correctly, and we excluded from analyses the few instances that L2 speakers subsequently revealed they had misinterpreted. Future work comparing how L1 and L2 speakers process *novel* metaphors may demonstrate greater convergence between the two groups. That is, we know of no reason to predict novel metaphors to be less emotionally engaging for native speakers than conventional metaphors, so amygdala activation in L1 speakers for novel metaphors is expected, at least as long as the novel metaphors are well-formed, apt, or in other words, *good* (Littlemore, Pérez-Sobrino, Houghton, Shi, & Winter, 2018). Moreover, to the extent that novel metaphors are more distinct from literal sentences (see also Bambini, Canal, Resta, & Grimaldi, 2019) than conventional metaphors, the contrast between novel > literal could reveal amygdala activation in an L2 group. However, there is some evidence that L2 speakers process novel and conventional metaphors similarly (Jankowiak et al., 2017; Mashal et al., 2015), so this requires further research.

4.1 Conclusion

The present study compares the processing of conventional metaphorical expressions in native speakers of German and in proficient German speakers whose first language is Italian. Results show that native speakers' process metaphors in ways that are consistent with previous findings: As Metaphoricity increased, there was more activity in the amygdala as well as in the IFG, ACC and adjacent cortices. At the whole brain level, the L2 speakers showed greater activation in the extended language network and part of the switching network in an indistinguishable range of areas for both literal and metaphorical sentences, when compared with native speakers. Thus we find the expected differences between L1 and L2 speakers' processing overall in the fronto-temporal language network. However, we do not find evidence that activation increases specifically with Metaphoricity in L2 speakers at the whole brain level or even in a region of interest analysis of the amygdala. The results suggest that L2 speakers may process conventional metaphors and literal paraphrases as more alike than native speakers do.

Authors' specific contributions

FMMC and AEG conceived and designed the experiment and wrote the manuscript together. FMMC designed more technical aspects of the experiment, set it up, programmed it, analyzed the data, and she and AEG interpreted the data. NM created the linguistic stimuli, validated them through questionnaires, and recruited and tested the majority of participants.

Acknowledgements

This work, including the acquisition of fMRI data, was fully funded by an Einstein Visiting Fellowship awarded AEG, in conjunction with the Cluster of Excellence "Languages of Emotion", Freie Universität Berlin. The data analysis was conducted at and therefore supported by Lancaster University. We thank Jeremie Güsten for the statistical selection of the linguistic materials used and for part of data collection, and Chun-Ting Hsu for his advice on some of the fMRI data analyses performed.

References

Abutalebi, J., Della Rosa, P. A., Ding, G., Weekes, B. S., Costa, A., & Green, D. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex, 49*, 905-911. doi:10.1016/j.cortex.2012.08.018

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*, 242-275. doi:10.1016/j.jneuroling.2006.10.003

Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *The Neuroscientist, 18*, 251-270. doi:10.1177/1073858411403316

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incogruity in bilinguals. *Brain and Language, 39*, 187-205.

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology, 7*, 559. doi:10.3389/fpsyg.2016.00559

Bambini, V., Canal, P., Resta, D., & Grimaldi, M. (2019). Time course and neurophysiological underpinnings of metaphor in literary context. *Discourse Processes, 56*, 77-97. doi:10.1080/0163853X.2017.1401876

Bambini, V., Ghio, M., Moro, A., & Schumacher, P. B. (2013). Differentiating among pragmatic uses of words through timed sensicality judgments. *Frontiers in Psychology, 4*, 938. doi:10.3389/fpsyg.2013.00938

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience, 4*, 417-422. doi:10.1093/scan/nsp053

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences, 16*, 240-250. doi:10.1016/j.tics.2012.03.001

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *International Review of Applied Linguistics in Language Teaching, 43*, 319-328.

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language - A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia, 50*, 2669-2683. doi:10.1016/j.neuropsychologia.2012.07.021

Bonin, P., Méot, A., & Bugaiska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods, 45*, 1259-1271. doi:10.3758/s13428-013-0331-4

Bookheimer, S. Y. (2002). Functional MRI of language: New approaches to understanding the cortical organisation of semantic processing. *Annual Review of Neuroscience, 25*, 151-188. doi:10.1146/annurev.neuro.25.112701.142946

Botvinick, M., Nystrom, L. E., Fissell, K., Cater, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179-181.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological Review, 112*, 193.

Bowes, A., & Katz, A. (2015). Metaphor creates intimacy and temporarily enhances theory of mind. *Memory & Cognition, 43*, 953-963. doi:10.3758/s13421-015-0508-4

Brabec, J., Rulseh, A., Hoyt, B., Vizek, M., Horinek, D., Hort, J., & Petrovicky, P. (2010). Volumetry of the human amygdala - an anatomical study. *Psychiatry Research, 182*, 67-72. doi:10.1016/j.pscychresns.2009.11.005

Briellmann, R. S., Saling, M. M., Connell, A. B., Waites, A. B., Abbott, D. F., & Jackson, G. D. (2004). A high-field functional MRI study of quadri-lingual subjects. *Brain and Language, 89*, 531-542.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1-38. doi:10.1196/annals.1440.011

Burgers, C., Konijn, E. A., Steen, G. J., & Iepsma, M. A. R. (2015). Making ads less complex, yet more creative and persuasive: the effects of conventional metaphors and irony in print advertising. *International Journal of Advertising*. doi:10.1080/02650487.2014.996200

Cameron, L. (2008). Metaphor and talk. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 197-211). New York: Cambridge University Press.

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage, 59*, 3212-3221. doi:10.1016/j.neuroimage.2011.11.079

Carston, R. (2012). Metaphor and the literal/nonliteral distinction. In K. Allan & K. Jaczszolt (Eds.), *Cambridge Handbook of Pragmatics* (pp. 469-492). Cambridge: Cambridge University Press.

Chang, C. T., & Yen, C. T. (2013). Missing ingredients in metaphor advertising: The right formula of metaphor type, product type, and need for cognition. *Journal of Advertising, 42*, 80-94.

Charteris-Black, J. (2011). *Politicians and rhetoric: The persuasive power of metaphor*: Springer.

Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *The Journal of Neuroscience, 19*, 3050-3056.

Cieślika, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22*, 115-144.

Cieślika, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language, 116*, 136-150. doi:10.1016/j.bandl.2010.09.007

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*, 2585-2595. doi:10.1162/jocn_a_00654

Citron, F. M. M., Güsten, J., Michaelis, N., & Goldberg, A. E. (2016b). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage, 139*, 218-230. doi:10.1016/j.neuroimage.2016.06.020

Citron, F. M. M., Lee, M., & Michaelis, N. (2019b). Affective and psycholinguistic norms for German conceptual metaphors (COMETA). *Behaviour Research Methods, Manuscript accepted pending minor revisions*.

Colbeck, K. L., & Bowers, J. S. (2012). Blinded by taboo words in L1 but not in L2. *Emotion, 12*, 217-222. doi:10.1037/a0026387

Columbus, G., Sheikh, N. A., Côté-Lecaldare, M., Häuser, K., Baum, S. R., & Titone, D. A. (2015). Individual differences in executive control relate to metaphor processing: an eye movement study of sentence reading. *Frontiers in Human Neuroscience, 8*, 1057. doi:10.3389/fnhum.2014.01057

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition, 30*, 958-968.

Crawford, L. E. (2009). Conceptual metaphors of affect. *Emotion Review, 1*, 129-139.

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science, 21*, 54-59. doi:10.1177/0963721411430832

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*, 109-114.

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development, 25*, 204-222. doi:10.1080/01434630408666529

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175-197. doi:10.1017/S1366728902003012

Fabbro, F. (2001). The bilingual brain: Cerebral representation of languages. *Brain and Language, 79*, 211-222. doi:10.1006/brln.2001.2481

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and symbolic activity, 2*, 239-250.

Ferstl, E. C. (2010). Neuroimaging of text comprehension: Where are we now? *Italian Journal of Linguistics, 22*, 61-88.

Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29*, 581-593.

Forgács, B., Bohrn, I. C., Baudewig, J., Hofmann, M. J., Csaba, P., & Jacobs, A. M. (2012). Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. *NeuroImage, 63*, 1432-1442. doi:10.1016/j.neuroimage.2012.07.029

Forgács, B., Lukács, Á., & Pléh, C. (2014). Lateralized processing of novel metaphors:

disentangling figurativeness and novelty. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2014.01.003

Friederici, A. D., Rüschemeyer, S. A., Hahne, A., & Fiebach, C. (2003). The role of the left inferior frontal and superior temporal cortex in sentence comprehension: Localising syntactic and semantic processes. *Cerebral Cortex, 13*, 170-177.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., & Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *NeuroReport, 12*, 2779-2783.

Gibbs, R. W. (2011). Evaluating conceptual metaphor theory. *Discourse Processes, 48*, 529-562.

Gibbs, R. W., Lima, P. L. C., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics, 36*, 1189-1210. doi:10.1016/j.pragma.2003.10.009

Giora, R. (1999). On the priority of salient meanings: studies of literal and figurative language. *Journal of Pragmatics, 31*, 919-929.

Glucksberg, S. (1998). Understanding metaphor. *Current Directions in Psychological Science, 7*, 39-43.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *NeuroReport, 13*, 15-19.

Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development, 25*, 223-247. doi:10.1080/01434630408666530

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics, 24*, 561-579. doi:10.1017.S0142716403000286

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 257-283). Clevedon, UK: Multilingual Matters.

Hayakawa, S., & Keysar, B. (2018). Using a foreign language reduces mental imagery. *Cognition, 173*, 8-15.

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science, 28*, 1387-1397.

Horton, W. S. (2007). Metaphor and reader's attribution of intimacy. *Memory & Cognition, 35*, 87-94.

Hsu, C.-T., Jacobs, A. M., & Conrad, M. (2015). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex, 63*, 282-295. doi:10.1016/j.cortex.2014.09.002

Jankowiak, K., Rataj, K., & Naskręcki, R. (2017). To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension. *PLoS ONE, 12*, e0175578. doi:10.1371/journal.pone.0175578

Kanske, P., & Kotz, S. A. (2011). Emotion triggers executive attention: Anterior cingulate cortex and amygdala responses to emotional words in a conflict task. *Human Brain Mapping, 32*, 198-208. doi:10.1002/hbm.21012

Kasparian, K. (2013). Hemispheric differences in figurative language processing: Contributions of neuroimaging methods and challenges in reconciling current empirical findings. *Journal of Neurolinguistics, 26*, 1-21. doi:10.1016/j.jneuroling.2012.07.001

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language, 28*, 375-385.

Koda, K. (1996). L2 word recognition research: A cricical review. *The Modern Language Journal, 80*, 450-460.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*, 416-430. doi:10.1016/j.actpsy.2008.02.001

Lai, V. T., & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language, 127*, 484-496. doi:10.1016/j.bandl.2013.09.010

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research, 1284*, 145-155. doi:10.1016/j.brainres.2009.05.088

Lai, V. T., van Dam, W., Conant, L. L., Binder, J. R., & Desai, R. H. (2015). Familiarity differentially affects right hemisphere contributions to processing metaphors and literals. *Frontiers in Human Neuroscience, 9*, 44. doi:10.3389/fnhum.2015.00044

Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica, 54*, 186-196.

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilingals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia, 50*, 1362-1371. doi:10.1016/j.neuropsychologia.2012.02.021

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition, 36*, 1103-1121. doi:10.3758/MC.36.6.1103

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*, 423-428. doi:10.1093/scan/nsp052

Littlemore, J., Chen, P. T., Koester, A., & Barnden, J. (2011). Difficulties in metaphor comprehension faced by international students whose first language is not English. *Applied Linguistics, 32*, 408-429. doi:10.1093/applin/amr009

Littlemore, J., & Low, G. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied Linguistics, 27*, 268-294. doi:10.1093/applin/aml004

Littlemore, J., Pérez-Sobrino, P., Houghton, D., Shi, J., & Winter, B. (2018). What makes a good metaphor? A cross-cultural study of computer-generated metaphor appreciation. *Metaphor* *and Symbol, 33*, 101-122. doi:10.1080/10926488.2018.1434944

Luk, G., Green, D., Abutalebi, J., & Grady, C. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes, 27*, 1479-1488. doi:10.1080/01690965.2011.613209

Maldjian, J. A., Laurienti, P. J., Kraf, R. A., & Burdette, J. B. (2003). An automated method for neuro-anatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage, 19*, 1233-1239.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62*, 103-134. doi:10.1146/annurev-psych-120709-145406

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*, 70-82. doi:10.1016/S0093-934X(02)00535-7

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*.

Martin, A. (1999). Automatic activation of the medial temporal lobe during encoding: Lateralized influences of meaning and novelty. *Hippocampus, 9*, 62-70. doi:10.1002/(SICI)1098-1063(1999)9:1<62::AID-HIPO7>3.0.CO;2-K

Mashal, N., Borodkin, K., Maliniak, O., & Faust, M. (2015). Hemispheric involvement in native and non-native comprehension of conventional metaphors. *Journal of Neurolinguistics, 35*, 96-108.

Mashal, N., & Faust, M. (2008). Right hemisphere sensitivity to novel metaphoric relations: Application of the signal detection theory. *Brain and Language, 104*, 103-112. doi:10.1016/j.bandl.2007.02.005

Mashal, N., Faust, M., & Hendler, T. (2005). The role of the right hemisphere in processing nonsalient metaphorical meanings: Application of principal components analysis to fMRI data. *Neuropsychologia, 43*, 2084-2100.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies. *Cortex, 44*, 848-860. doi:10.1016/j.cortex.2007.04.004

McNab, F., Leroux, G., Strand, F., Thorell, L., Bergman, S., & Klingberg, T. (2008). Common and unique components of inhibition and working memory: An fMRI, within-subjects investigation. *Neuropsychologia, 46*, 2668-2682. doi:10.1016/j.neuropsychologia.2008.04.023

McQuarrie, E. F., & Mick, D. G. (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research, 22*, 424-438. doi:10.1086/209459

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology, 47*, 36-45.

Ortony, A., & Fainsilber, L. (1987). *The role of metaphors in descriptions of emotions.* Paper presented at the 1987 workshop on theoretical issues in natural language processing.

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *NeuroImage, 21*, 623-631. doi:10.1016/j.neuroimage.2003.09.069

Parker Jones, O., Green, D., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., . . . Price, C. J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex, 22*, 892-902. doi:10.1093/cercor/bhr161

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*, 405-428. doi:10.1080/00207594.2012.743665

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology, 15*, 202-206. doi:10.1016/j.conb.2005.03.007

Perani, D., Paulesu, E., Sebastian-Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., . . . Mehler, J. (1998). The bilingual brain. Proficency and age of acquisition of the second language. *Brain, 121*, 1841-1852.

Pollio, H. R., Barlow, J. M., Fine, H. J., & Pollio, M. R. (1977). *Psychology and the poetics of growth: Figurative language in psychology, psychotherapy, and education.*: Erlbaum.

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*, 600-610. doi:10.1016/j.neuroimage.2012.06.022

Rojo, A., Ramos, M., & Valenzuela, J. (2014). The emotional impact of translation: A heart rate study. *Journal of Pragmatics, 71*, 31-44. doi:10.1016/j.pragma.2014.07.006

Rüschemeyer, S. A., Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *NeuroImage, 31*, 354-365. doi:10.1016/j.neuroimage.2005.11.047

Schmidt, G. L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond laterality: A critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society, 16*, 1-5. doi:10.1017/S1355617709990543

Seeley, W. W., Menon, V., Schatzberg, A., F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*, 2349-2356.

Semino, E. (2008). *Metaphor in discourse*. Cambridge: Cambridge University Press.

Siyanova-Chanturia, A., Canal, P., & Heredia, R. R. (2019). Event-related potentials in monolingual and bilingual non-literal language processing. In J. W. Schwieter (Ed.), *The handbook of the neuroscience of multilingualism* (pp. 508-529): Wiley Blackwell.

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Sciences*.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the USA, 104*, 12530-12535. doi:10.1073/pnas.0609927104

Türker, E. (2016). The role of L1 conceptual and linguistic knowledge and frequency in the acquisition of L2 metaphorical expressions. *Second Language Research, 32*, 25-48. doi:10.1177/0267658315593336

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., . . . Sadato, N. (2012). Distinction between the literal and intended meanings of sentences: A functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex, 48*, 563-583. doi:10.1016/j.cortex.2011.01.004

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. doi:10.3389/fnhum.2014.00583

Winter, B. (2016). Taste and smell words form and affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2016.1193619

Yang, J. (2014). The role of the right hemisphere in metaphor comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Human Brain Mapping, 35*, 107-122. doi:10.1002/hbm.22160

Appendix A. Clusters showing significant BOLD signal change at the whole-brain level for the main effect of Group, in both directions: L2 > L1 speakers; L1 > L2 speakers. At the voxel level, a significance threshold of $p < 0.005$ was applied, followed by FDR correction at the cluster level ($p < 0.05$). Legend: Hemi. = hemisphere, L = left, R = right; cluster size is in voxels, T = peak t value; X, Y, Z = MNI stereotactic space coordinates.

Appendix A.

Broader area	Hemi.	Region	Cluster size	T	X, Y, Z
L2 > L1					
Fronto-temporal lobes	L	Inferior frontal gyrus, pars opercularis (BA 45)	134	4.52	-48 15 18
		Rolandic operculum		4.24	-60 9 3
		Anterior superior temporal lobe		3.64	-57 -6 -6
Pericentral lobes	L	Pre-central gyrus	652	5.93	-18 -21 72
		Pre-cental gyrus (BA 6)		5.32	-48 -12 57
		Pre-cental gyrus (BA 6)		5.11	-48 -3 48
	R	Pre-cental gyrus (BA 6)	107	4.16	42 -21 63
		Pre-central gyrus		3.88	33 -15 60
		Pre-cental gyyrus (BA 6)		3.69	39 -9 45
Temporal lobe	L	Middle temporal gyrus	130	4.94	-57 -24 0
		Middle temporal gyrus		4.54	-51 -42 9
Medial temporal lobe	R	Hippocampus	69	4.83	30 -36 -3
		Parahippocampal gyrus		4.02	18 -36 3
		--		3.92	27 -39 9
	R	--	78	4.14	15 -15 -18
		Hippocampus		4.14	30 -15 -18
Basal ganglia	R	Caudate nucleus (head)	487	6.27	18 24 6
		--		5.55	0 -18 12
		Caudate nucleus (head)		4.79	9 18 6
	L	--	161	5.60	-18 -12 -9
		Putamen		4.75	-18 9 -6
		Putamen		4.70	-18 9 3
Occipito-temporal lobes	L	Inferior temporal lobe	273	5.32	-51 -57 -12
		--		5.31	-42 -45 -12
		Cerebellum		3.94	-45 -66 -27
	R	Calcarine fissure	109	4.93	15 -99 -6
		Middle occipital gyrus (BA 18)		4.68	33 -96 0
		Inferior occipital gyrus		4.28	33 -87 -3
Cerebellum	R	Cerebellum	325	4.65	6 -78 -18
		Inferior occipital gyrus		4.61	45 -66 -15
		Cerebellum		4.50	48 -63 -27
L1 > L2					
Frontal lobe	R	Middle frontal gyrus (BA 9)	95	4.72	27 30 33
		Middle frontla gyrus		3.56	27 42 33
		Middle frontal gyrus (BA 9)		3.50	36 15 36
Temporal lobe	L	Superior temporal gyrus	170	5.38	-42 -30 6
		Insula		5.35	-39 -12 3
		Superior temporal gyrus		4.33	-48 -24 9
	R	Superior temporal gyrus	204	5.85	57 -12 6
		Heschl's gyrus		5.00	45 -21 9
	R	Middle temporal gyrus	101	4.44	63 -57 3
		Middle temporal gyrus		4.01	57 -66 3
		Middle temporal gyrus		3.83	57 -57 15
Occipital lobe	R	Cuneus (BA 19)	2580	6.96	9 -87 24
		Pre-cuneus		6.68	9 -66 27
		Lingual gyrus (BA 30)		6.21	6 -69 3
	L	Cerebellum	114	4.41	-9 -48 -21
		Lingual gyrus		3.75	-15 -48 -9
		Lingual gyrus		3.50	-21 -45 -3