You say *potato*, I say *tǔdòu*: How speakers of different languages share the same concept

Benjamin D. Zinszer (bzinszer@mail.bcs.rochester.edu)
Andrew J. Anderson (aanderson@mail.bcs.rochester.edu)
Rochester Center for Brain Imaging
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627 USA

Olivia Kang (olivia.e.kang@dartmouth.edu)
Thalia Wheatley (thalia.wheatley@dartmouth.edu)
Department of Psychological and Brain Sciences
Dartmouth College
Hanover, NH 03755 USA

Rajeev Raizada (rajeev.raizada@gmail.com)
Rochester Center for Brain Imaging
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627 USA

Abstract

When a speaker of English and a speaker of Chinese think about the same object, their brains are representing a shared concept. However, we don't know how similarity in the concepts evoked by words is manifested in the brains of speakers of different languages. We have previously shown that neural similarity relations are strongly conserved across subjects, allowing across-subject decoding (Raizada & Connolly, 2012). Here we extend that result to translating word-elicited activations across groups of speakers of Chinese and English. Specifically, by matching the neural similarity relations elicited by a set of seven Chinese words, presented to Chinese speakers, with the neural similarities elicited by the equivalent English words presented to English speakers, we are able to translate between the English and Chinese words with 100% accuracy, based only on the patterns of functional activity that they elicit. This demonstration provides evidence for the conservation of semantic relations between concepts across different languages.

Keywords: MVPA, neural decoding, semantics, conceptual representation

Introduction

Recent innovations in multi-voxel pattern analyses (MVPA) and machine learning have enabled cognitive neuroscientists to predict patterns of brain activity for word stimuli by generalizing from a training set of other words and their associated functional neuroimaging data. This predictive power enables accurate re-association of observed brain activity with the specific stimuli that are most likely to have elicited that activity, an inferential procedure known as neural decoding. Neural decoding has allowed generalization to novel words (e.g., Mitchell *et al.*, 2008) and across participants (e.g., Raizada & Connolly, 2012).

Recent studies of bilingual speakers (Buchweitz *et al.*, 2012; Correia *et al.*, 2014, 2015) have shown that within a single bilingual person's brain, there are decodable associations between mental representations of translation equivalent words across the two languages. However, such decoding may simply detect within-subject associative pairings rather than relations between the semantic structures of the two languages. Here we ask how semantic relations between word-elicited concepts are conserved across different languages and whether this relationship is reflected in the neural representations across speakers of different languages. These representations could then be used to inform inferences about semantic similarity structures in two languages.

Neural Translation

To date, neural translation has only been demonstrated within bilingual participants, associating an individual's neural representations for words in one language with translation equivalents in the other language. This approach links the bilingual's languages at the level of semantic or conceptual representation, but embodiment hypotheses (Hauk, 2006; Binder & Desai, 2011) suggest that these representations should be comparable across speakers of different languages based on associations with perceptual experiences. For example, the appearance, sound, and general functions of a dog would be roughly the same for speakers of any language.

Achieving neural translation across speakers of different languages requires sufficient commonality in conceptual representations to associate them across languages. However, research in lexical semantics and translation has repeatedly demonstrated translation ambiguity between

languages, even for concrete nouns (Degani & Tokowicz, 2010; Malt & Majid, 2013). Successful neural translation across speakers of different languages would shed light on how linguistic representations are conceptually grounded despite language-specific variations in meaning.

One limitation of the previous studies of neural translation is that bilinguals' semantic representations are likely to rely on a shared conceptual store for both languages (see Dong *et al.*, 2005; and Ameel *et al.*, 2009). Behavioral studies of bilinguals demonstrate that these semantic representations reflect the mutual influence of first and second language norms (Dong *et al.*, 2005; Zinszer *et al.*, 2014), suggesting that within-subject neural translation draws on an individual's cross-language conceptual stores for both languages. In this light, neural translation *within* a bilingual may not greatly differ from the task of decoding monolinguals' neural activity.

Localization and Embodiment

Specific anatomical regions such as the fusiform cortex and parahippocampal gyrii have been identified as hubs for the synthesis of perceptual features in conceptual representations (Barsalou, 2008; Martin, 2007). Neural decoding studies have corroborated this claim by demonstrating that individual participants' multi-voxel patterns of activity in the ventral temporal cortex can be decoded based on observations from other speakers (e.g., Raizada & Connolly, 2012).

The previous decoding studies within bilinguals have similarly identified specific cortical regions for which multi-voxel responses generalize across languages and allow decoding of one language's activation patterns based on the other language (Buchweitz *et al.*, 2012; Correia *et al.*, 2014). Voxels showing cross-language correlation in bilinguals were widely distributed in these studies, but consistent with Barsalou (2008) and Martin's (2007) accounts, the parahippocampal gyrii were among the regions of cross-language stability (Buchweitz *et al.*, 2012).

The Present Study

In this study, we extend neural translation (neural decoding of words across languages) to independent groups of participants for each language. We use MVPA to compare distributed functional brain activity for speakers of Chinese and English reading words in their respective native language. We ask whether the similarity structures for neural representations of word-elicited concepts are sufficiently similar across languages to perform neural decoding on group-level data and translate words in one language into the other language.

Method

Participants

Eleven native speakers of English (4 M / 7 F) and eleven native speakers of Mandarin Chinese (3 M / 8 F) were recruited at Dartmouth College. All participants were undergraduate students, graduate students, or post-doctoral researchers. Participants self-reported being native speakers of English or Mandarin Chinese, defined as being born in their native language environment and speaking that language as their earliest language.

Materials

We selected seven translation equivalent words in English and Chinese before the study, meeting four criteria: (1) concrete nouns (2) monosyllabic in both languages, (3) represented by a single Chinese character, and (4) unlikely for English translations to be known by the Chinese participants (see Table 1 for list). To insure that criterion (4) was met, Chinese participants who accurately translated more than two of the critical stimuli to English in a post-scan quiz were excluded from analysis. The critical stimuli were presented in three different font faces (English: Helvetica, American Typewriter, and Times New Roman; Chinese: STFangSong, Kai, and STSong) to reduce the influence of visual similarity on neural representations of the stimuli. The functional activity elicited by these words forms the basis of all the analyses presented here.

Table 1. Critical stimuli in English and Chinese

English	Chinese (pinyin)
axe	斧 (fǔ)
broom	帚 (zhou)
gown	袍 (páo)
hoof	蹄 (tí)
jaw	颚 (è)
mule	骡 (luó)
raft	筏 (fá)

Participants completed a semantic relatedness task involving catch trials and filler words interspersed between the seven critical stimuli to encourage them to think about word meanings. Filler words were not used in any of the fMRI analyses. Of the 42 filler words (translation equivalents in both languages, did not need to meet the critical criteria), half were semantically related to one of the critical stimuli (e.g., axe – log) and half were semantically unrelated (e.g., axe – moth) for a total of three related words and three unrelated words for each critical stimulus.

Stimuli for this task were presented as black text on a gray background via projector to a screen behind the MRI scanner. Participants viewed the projected words through a mirror attached to the scanner's head coil.

Procedure

Experimental procedures were approved by the Dartmouth Committee for the Protection of Human Subjects. Participants completed the semantic relatedness task while undergoing functional magnetic resonance imaging (fMRI). Words were presented for 1750 ms, followed by a 5750 ms

fixation cross. If a catch-word was presented in red text with a "?" (e.g., "moth?"), participants responded by indicating whether the catch-word was semantically related to the word immediately preceding. Catch-words were always filler words and occurred in approximately one third of trials to encourage participants to think about the meanings of each stimulus word. Each functional run was composed of 45 to 50 stimulus presentations, about seven minutes in duration. Participants completed seven functional runs for a total of 35 presentations per critical stimulus word.

Image Acquisition & Processing

Scanning Parameters The study was performed using a Philips Intera Achieva 3-Tesla scanner (Philips Medical Systems, Bothell, WA) with a SENSE (SENSEitivity Encoding) 32-channel head coil. Anatomical images were collected using a high-resolution 3D Magnetization-prepared rapid gradient echo sequence (220 slices, 1mm isotropic voxels, FOV=240mm, acquisition matrix= 256x256). Functional images were collected in 7 runs using echo planar functional images sensitive to blood oxygenation level-dependent (BOLD) contrast (TR= 2000ms, TE=35ms, flip angle=90 degrees, 3 mm in-plane resolution). During each of the functional runs, 175 sets of axial images (35 slices/volume) were collected in an interleaved fashion across the whole brain.

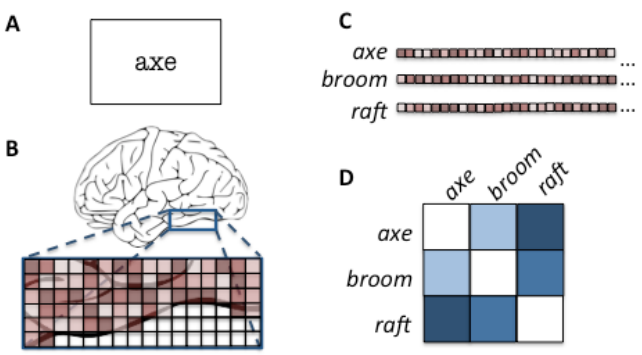

Figure 1. Procedure for computing a neural similarity matrix: (A) Stimulus is presented during functional imaging. (B) Individual voxel responses to stimulus are measured or estimated. (C) Responses for each stimulus are compared as 1 x *n* vectors for *n* voxels. (D) Stimulus representations are correlated to generate the neural similarity matrix.

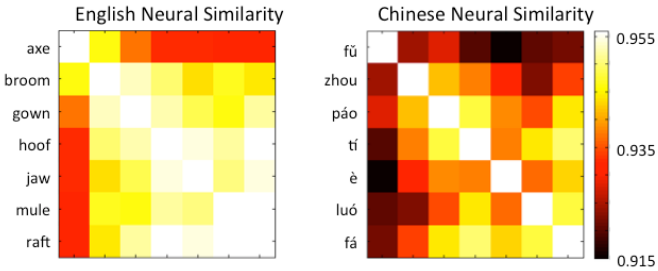

Figure 2. Neural similarity matrices for each language group.

Pre-processing and Estimation Functional images across seven runs were realigned to the mean image and resliced. A general linear model was estimated with separate regressors for each of the seven critical stimuli and a regressor for the response type (catch-trial or none). Separate parameters were estimated for each functional run, and then averaged in contrasts defined for each of the critical stimuli.

MVPA and Neural similarity Individual participants' multi-voxel patterns for the critical stimuli were computed separately in 96 anatomical ROIs (48 in each hemisphere), as defined by the Harvard-Oxford Atlas (http://www.fmrib.ox.ac.uk/fsl/).

Figure 1 illustrates the procedure for calculating neural similarity in a single participant. The response pattern for each critical stimulus (a word) was defined by the contrast map (beta weights) estimated for it in the first level model. The response patterns to each of the seven critical stimuli were then correlated, resulting in a 7-by-7 neural similarity matrix in which each stimulus is described by the Pearson correlation of its functional response pattern to that of the other six stimuli.

Results

Behavioral Responses

Catch-word trials were checked for response rate and response time to be sure that participants were reading the words. Mean response rate was 83% (*SD*=20%) and mean RT was 1398 ms (*SD*=167 ms). No measure of response accuracy was performed because the semantic relatedness judgments are subjective.

Neural Similarity

Individual neural similarity matrices were computed for each participant based on their unique patterns of functional activity for the seven critical stimuli. The similarity matrices were transformed using Fisher's *r*-to-*z* (the inverse hyperbolic tangent) to normalize the *r* distribution, and a group similarity matrix was computed by averaging individual participants' matrices for each language. The whole-brain similarity matrices for English and Chinese are illustrated at left (Figure 2).

Similarities between the Chinese and English whole-brain neural similarity structures were reflected in a Pearson correlation between the unique values in each matrix (the left triangle, excluding diagonals). Chinese and English were strongly correlated, r=0.89, p<0.001. This cross-language correlation was also computed within each ROI of the Harvard-Oxford atlas. Table 2 (next page) lists the top twelve ROIs by the magnitude of their correlation. Many ROIs showed very strong correlation between languages (79 significant at p<0.05 level, 41 significant after Bonferroni correction), particularly in bilateral temporal and parietal areas (see Figure 3 on the next page, visualized using the xjView toolbox available at http://alivelearn.net/xjview).

Neural Translation

The English and Chinese group neural similarity data in each ROI were used to attempt neural decoding of one language using the neural similarity patterns obtained for the other language. This between-groups decoding provides a neurally grounded form of translation wherein Chinese words can be matched to English words based only on their respective brain representations, via the neural similarity structures for each language.

To achieve neural translation, a reference matrix (e.g., the English group neural similarity) is compared to every possible permutation of stimuli in the test matrix (e.g., the Chinese group neural similarity). If the neural similarity structures are similar enough between two languages, the permutation of the test matrix most highly correlated with the reference matrix will be the correct set of translations. A threshold for statistical significance was computed by taking the 95th percentile of the full distribution of accuracy scores for all possible permutations. The 95th percentile of the accuracy distribution for all permutations was 0.4286. Thus scores above this threshold have a 0.05 probability of occurring by random selection (see Raizada & Connolly, 2012 regarding permutation testing). Bonferonni correction for multiple comparisons (96 ROIs) results in a significance threshold of 0.7143.

All ROIs that achieved 100% decoding accuracy between languages are included and highlighted in Table 2. Switching the reference and test matrices yields identical results. Many ROIs, however, achieved accuracy scores that were significantly above chance before correcting for multiple comparisons. Figure 4 (visualized using xjView) illustrates decoding accuracy across a sample of cortical regions. The whole-brain similarity structures yielded a decoding accuracy of 0.7143.

While we found several cortical regions that were strongly correlated between languages, only a few of these regions resulted in accurate cross-language decoding. In the left hemisphere, the anterior parahippocampal (r=0.74) and postcentral gyrii (r=0.89) produced the best decoding results. In the right hemisphere, the frontal orbital cortex (r=0.78), anterior cingulate gyrus (r=0.76), anterior supramarginal gyrus (r=0.87), and posterior inferior temporal gyrus (r=0.92) also produced decoding scores of 1.0.

Cross-language correlation was a strong predictor of this decoding accuracy (r=0.69 across 96 ROIs), however some regions that correlated highly between languages were not successful for decoding: left central opercular cortex (r=0.87, Acc=0.29), right middle frontal gyrus (r=0.83, Acc=0.29), left posterior supramarginal gyrus (r=0.80, Acc=0.29), and the temporo-occipital division of the left middle temporal gyrus (r=0.79, Acc=0.29). In these regions, even higher correlations were obtained for incorrect permutations of the words, leading to lower neural translation accuracies.

Table 2. Cross-language correlation and decoding accuracy in select ROIs from the Harvard-Oxford brain atlas.

HO ROI	Anatomical Region	*r*	Acc.
18	L Supramarginal Gyrus, anterior	0.93	0.71
14	R Inferior Temporal Gyrus, posterior	0.92	0.57
16	R Postcentral Gyrus	0.92	1.00
06	L Precentral Gyrus	0.90	0.43
16	L Postcentral Gyrus	0.89	1.00
18	R Supramarginal Gyrus, anterior	0.87	1.00
15	R Inferior Temporal Gyrus, temporooccipital part	0.87	0.71
11	L Middle Temporal Gyrus, posterior	0.87	0.29
41	L Central Opercular Cortex	0.87	0.71
06	R Precentral Gyrus	0.86	0.57
37	L Temporal Fusiform Cortex, posterior	0.86	0.71
12	R Middle Temporal Gyrus, temporooccipital part	0.86	0.57
28	R Cingulate Gyrus, anterior	0.78	1.00
32	R Frontal Orbital Cortex	0.76	1.00
33	L Parahippocampal Gyrus, anterior	0.74	1.00

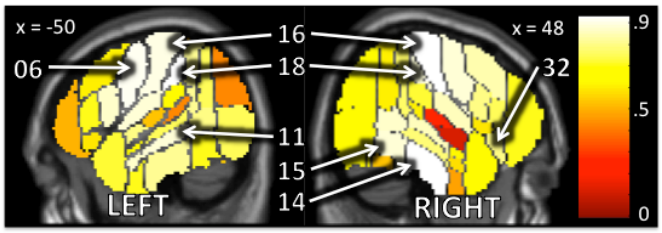

Figure 3. Cross-language correlation of neural similarity matrices projected onto cortical surface. See Table 2 for selected values.

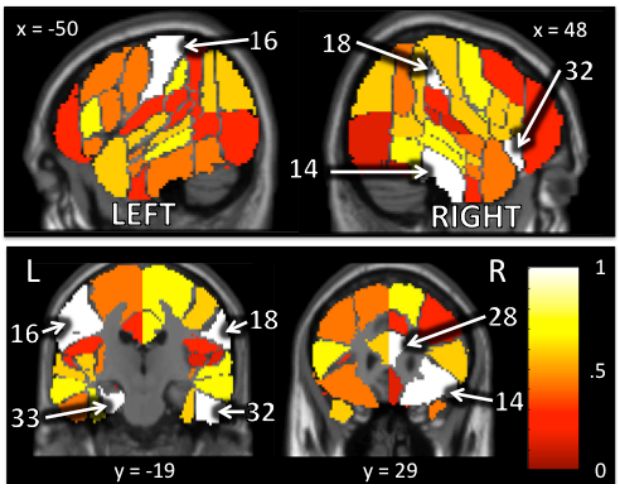

Figure 4. Decoding accuracy projected onto the cortical surface. See Table 3 for list of ROIs with 100% accuracy.

Discussion

In this study, we demonstrated that neural similarity representations for words in native speakers of Chinese and English are similar enough between languages to allow cross-language decoding or neural translation of seven words. Previous studies examining the representations of words across languages have compared patterns within bilingual speakers (Buchweitz *et al.*, 2012; Correia *et al.*, 2014, 2015), but these studies leave open the possibility that individual bilinguals represent two languages based on internally consistent but not generalizable grounds. By comparing across native speakers of each language, we greatly reduce the plausibility of this explanation and offer neurocognitive evidence for a grounded representation common across speakers and languages. Our demonstration also illustrates the possibility of achieving neurally informed translation in the future based on the relative similarity of native speakers' neural responses to words in each language.

Localization of Effects

Left hemisphere regions producing the best decoding accuracy in this study were consistent with cross-language stability findings in the previous studies. Left postcentral gyrus (Buchweitz *et al.*, 2012 and Correia *et al.*, 2014) and left parahippocampal gyrus (Buchweitz *et al.*, 2012) both previously exhibited stability across languages for bilingual speakers. These regions have also been linked to processing of concepts related to tools and shelter (respectively, see Just *et al.*, 2010) and cross-modal semantic integration (Hickok & Poeppel, 2007).

Our results support the latter claim that postcentral and parahippocampal gyrii integrate semantic information from non-linguistic modalities. Given the particular roles of the postcentral gyrus and parahippocampal regions in somatosensory representation and memory retrieval, respectively, we also find support for the broader claim that these conceptual representations are grounded in multimodal somatosensory and episodic memories.

Several right hemisphere regions also provided high decoding accuracies. Although the left inferior temporal gyrus and left supramarginal gyrus have previously been implicated in semantically-based neural decoding (Raizada & Connolly, 2012; Correia *et al.*, 2014), we found that decoding accuracy was higher in the right hemisphere analogs of these structures.

Previous neural decoding studies have not specifically investigated right lateralization effects, but some explanation might be drawn from research on lateralization in semantic processing. Semantic information in the right hemisphere has long been hypothesized to represent coarser, message-level semantic representations (Beeman, 1993) and more recently been associated with processing semantically distant or novel associations and semantic context (Jung-Beeman, 2005; Vigneau *et al.*, 2011) such as in metaphor comprehension (Schmidt, DeBuse, & Seger, 2005; Vigneau *et al.*, 2011). In the present study, coarser representations may offer better cross-language symmetry than fine grained language- or culturally-specific information. Particularly since our critical stimuli were composed of only seven relatively distant concepts, the coarse representations for these concepts could be more consistent across languages than their left-lateralized, finer grained representations (such as the exact shape and appearance of a prototypical *broom* or *raft*).

Importantly this right hemisphere advantage for neural translation in the present study is observed between language groups. By contrast, Correia and colleagues' (2014) within-bilingual study of neural translation produced a relatively balanced set of left and right hemisphere regions that were stable across languages. However, in their study, left hemisphere generalization across languages may be supported by within-subject stability, drawing on bilinguals' semantic convergence (Dong *et al.*, 2005; Zinszer *et al.*, 2014).

Our results also identified the right anterior cingulate cortex for high decoding accuracy, which has typically been described as providing conflict monitoring for cognitive control (Botvinick, Cohen, & Carter, 2004; Shenav, Botvinick, & Cohen, 2013), including in the case language conflict in bilingualism (Abutalebi & Green, 2007; Green & Abutalebi, 2013). Concept-specific representation has not previously been demonstrated in the anterior cingulate, but we see some evidence of this specificity in our results. The right lateralization of this anterior cingulate effect may not be especially important, given the region's medial location. Brain normalization across participants may have been insufficiently precise to distinguish between left and right lateralized functions in the anterior cingulate. The left side also produced strong cross-language correlation (0.75) and above chance decoding accuracy before correction for multiple comparisons (0.57).

Translation by Neural Similarity

Our comparison of neural similarity structures in native speakers of Chinese and English yielded a successful translation between English and Chinese words based on the functional brain responses of separate groups of participants using each language. This ability to compare brain representations of words between speakers of different languages presents a new way of studying translation asymmetry, such as between abstract nouns for which experimental evidence indicates translation costs due to ambiguity (see Van Hell & De Groot, 1998). Neurally informed translation permits comparison of multiple translation candidates for their relative fitness to brain responses elicited by the other language. Further, language-specific and language-independent elements of brain representation can be contrasted by examining translation pairs for correlation to non-linguistic measures (e.g., visual

object information) and linguistic measures (e.g., word co-occurrence).

Our neural translation task was limited to a lexicon of seven words in each language and tested by selecting the permutation of words in one language that best approximated the neural similarity of words in the target language. While the permutation method worked well in the present study, it is computationally infeasible for even slightly larger lexicons, as the number of permutations that must be compared expands factorially (e.g., seven words have 5040 permutations, but ten words have over 3.6 million permutations). However, search optimization strategies offer a number of opportunities to refine the existing algorithm, which would allow neural translation to scale up to much larger lexicons.

Conclusion

In this study, we successfully extended across-participant neural decoding to groups of participants using different languages. Doing so, we identified semantic representations that are preserved across languages in the form of neural similarity structures. The distribution of these cross-language similarities across the cortex was consistent with previously identified regions (parahippocampal and postcentral gyrii) and implicated several right hemisphere structures informative to cross-language semantics.

References

Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. doi:10.1016/j.jneuroling.2006.10.003

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–45. doi:10.1146/annurev.psych.59.103006.093639

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, *44*, 80–120. doi:10.1006/brln.1993.1006

Binder, J. R., & Desai, R. H. R. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527-36. doi:10.1016/j.tics.2011.10.001

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*(12), 539-46. doi:10.1016/j.tics.2004.10.003

Buchweitz, A., Shinkareva, S. V, Mason, R., Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language*, *120*(3), 282-9. doi:10.1016/j.bandl.2011.09.003

Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience*, *34*(1), 332–8. doi:10.1523/JNEUROSCI.1302-13.2014

Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., & Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Frontiers in Psychology*. doi:10.3389/fpsyg.2015.00071

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. doi:10.1080/20445911.2013.796377

Hauk, O., Johnsrude, I., & Pulvermu, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(May), 393–402.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. doi:10.1016/j.tics.2005.09.009

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. doi:10.1146/annurev.psych.57.102904.190143

Raizada, R., & Connolly, A. (2012). What makes different people's representations alike: Neural similarity space solves the problem of across-subject fMRI decoding. *Journal of Cognitive Neuroscience*, *24*(4), 868–877.

Schmidt, G. L., DeBuse, C. J., & Seger, C. a. (2007). Right hemisphere metaphor processing? Characterizing the lateralization of semantic processes. *Brain and Language*, *100*, 127–141. doi:10.1016/j.bandl.2005.03.002

Shenhav, A., Straccia, M. a, Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, *17*(9), 1249–1254. doi:10.1038/nn.3771

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. doi:10.1017/S1366728998000352

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*(1), 577–593. doi:10.1016/j.neuroimage.2010.07.036

Zinszer, B. D., Malt, B. C., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in psychology, 5*(October), 1203. doi:10.3389/fpsyg.2014.01203