http://dx.doi.org/10.1016/j.bandl.2017.09.007 AUTHOR PREPRINT

Brain and Language

Commonalities and differences in the neural representations of English, Portuguese, and Mandarin sentences: when knowledge of the brain-language mappings for two languages is better than one

Ying Yang[a], Jing Wang[a], Cyntia Bailer[b], Vladimir Cherkassky[a], Marcel Adam Just[a]

[a] *Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA*
[b] *Department of Foreign Language and Literature, Federal University of Santa Catarina, Brazil*

Keywords:
Cross-language sentence decoding
fMRI concept signatures
Cross-language neural commonalities
Cross-language neural differences

ABSTRACT

This study extended cross-language semantic decoding (based on a concept's fMRI signature) to the decoding of sentences across three different languages (English, Portuguese and Mandarin). A classifier was trained on either the mapping between words and activation patterns in one language or the mappings in two languages (using an equivalent amount of training data), and then tested on its ability to decode the semantic content of a third language. The model trained on two languages was reliably more accurate than a classifier trained on one language for all three pairs of languages. This two-language advantage was selective to abstract concept domains such as social interactions and mental activity. Representational Similarity Analyses (RSA) of the inter-sentence neural similarities resulted in similar clustering of sentences in all the three languages, indicating a shared neural concept space among languages. These findings identify semantic domains that are common across these three languages versus those that are more language or culture-specific..

1. Introduction

Does the particular language we speak influence the way we represent concepts in our brains? This question has fascinated linguists, cognitive psychologists, neuroscientists and laymen for decades (Lucy 1997). Some researchers have argued that concept representation and cognitive processes are largely uninfluenced by language (e.g. Strømnes, 1974; Wierzbicka, 1992), while others have proposed that concept development is shaped by each language (e.g. Berman and Slobin, 2013; Johansson and Salminen, 1996). With the advent of cross-language/cross-cultural functional neuroimaging studies, the majority of concept domains have been shown to be represented in similar brain regions across languages and cultures (Enfield, 2015; Han & Northoff, 2008), although the neural encoding of a few perceptual and social domains (e.g. time representation, Boroditsky, 2001; historical and autobiographical events, Wang & Conway, 2004) may be influenced by language or culture (Boroditsky 2001; Levinson et al. 2002; Ozgen and Davies 2002; Cook et al. 2006; Athanasopoulos and Kasai 2008).

A recent eye-tracking study found that speakers of three orthographically distinct languages (Chinese, English and Finnish) manifested a great amount of reading behavior similarities (Liversedge et al. 2016), consistent with a commonality in concept representation in text reading despite graphemic and linguistic variation across languages.

The commonality has also been supported by several fMRI cross-language decoding studies: between Portuguese and English (Buchweitz et al., 2012; Yang et al., 2016), Dutch and English (Correia et al. 2014), and Mandarin and English (Zinszer et al. 2012). These studies demonstrated that similar concepts (referred to by translation equivalent words) induce similar neural activation patterns in speakers of different languages. Therefore, a machine learning algorithm can be trained to associate concepts with the neural activation patterns evoked by words or sentences in one language (the training language), and it can then recognize the neural activation pattern of the translation equivalent words in another language (the test language).

Previous demonstrations of commonality across different languages compared reading behaviors across three different languages, and cross-language neural decoding between a *pair* of languages. A missing link in developing a comprehensive theory of meta-language concept representation lies in cross-language neural decoding studies in ***three*** different languages. Investigating neural decoding across three languages is not a mere quantitative increase; it enables a scientific advance. Any pair of languages will have some degree of commonality, demonstrated when a classifier trained on the mapping between concepts and neural representations in one language can decode the concepts from their neural representations in another language. By contrast, training a classifier on two languages permits a test for the presence of language-specific mappings between concepts and neural representations. If such language-specific mappings exist, then there should be an advantage in training a classifier on two languages. A classifier trained on data from two languages has more opportunities to develop a mapping for items in the test language that are not universal but do occur in some other languages. The greater the number of training languages, the more likely that such non-universal items will occur in the classifier's training.

This study thus investigates whether a classifier trained on data from two languages will more accurately decode a third language than a classifier trained on an equivalent amount of data from only one language. If such a two-language advantage should emerge, it should further be possible to determine which semantic domains benefit most from the two-language training.

Furthermore, we asked these questions as they pertain to the neural representations of concepts as they occur in sentences rather than concepts in isolation. Sentences convey multi-concept complex messages that can describe an event or a status. These messages convey semantic information that transcends single word-level concepts. This study asked such questions of sentences in three languages, English, Portuguese and Mandarin, as they were read by native speakers of each language.

The commonalities in neural representations across different participants have been investigated in several previous multivariate pattern analyses (MVPA) of fMRI studies, such as comparing L1 and L2 representations in late bilinguals (Hsu, Jacobs, & Conrad, 2015); cross-language decoding at the single word level (Correia, et al. ,2014; Zinszer, Anderson, Kang, Wheatly, & Raizada, 2015), and sentence thematic information encoding in English (Frankland & Greene, 2014). This approach enables a comparison across people and languages of a brain activation pattern consisting of the activation levels of a set of voxels that are topographically distributed across multiple brain regions.

In the current study, a classifier was trained on the mapping between sentences and their activation patterns in one set of data, and then tested on an *independent* set of data. Three situations were compared: a classifier was trained on the mapping in two languages and then it was tested on a third language; a classifier was trained on the mapping in one language and then it was tested on another language; a classifier was trained on the mapping in one language and then it was tested on the same language. The amount of training data was equated in the three cases. These three situations and the languages involved are shown in Table 1.

1.1 Hypotheses

Four hypotheses were tested. First, classifiers trained on two languages (e.g. English and Portuguese) should generalize better to (classify sentences more accurately in) a third language (e.g. Mandarin) than classifiers trained on either one of the two training languages alone (e.g. English) when the amount of training data is equated. Furthermore, the classification accuracy resulting from training a classifier on two other languages should be more similar to the within-language accuracy than the accuracy resulting from training a classifier on only one other language.

Second, we hypothesize that any such *two-language advantage* (i.e. the accuracy boost from training on two languages compared to training on one language) will be greater for concept domains that are more language- or culture-specific than concept domains that are language- or culture-general, because the latter won't derive additional benefit from a mapping in a second training language. For example, words naming social interactions such as *marriage* may be more culture-specific than words naming physical objects such as *apple;* so the decoding of a concept like *marriage* may show more of a two-language advantage than the decoding of a concept like *apple*. Third, the meta-language neural commonality should not be affected by the distances between the superficial structures of the languages. For instance, the meta-language concept representations may not be more different between English and Mandarin than between English and Portuguese despite only the latter two being Indo-European. Fourth, the within-language inter-sentence neural similarity pattern, as computed by representational similarity analysis (RSA), should show commonalities across all three languages, indicating that the semantic space and semantic relationships among sentences is similar across languages.

2. Materials and Methods

2.1 Participants

Three groups of participants were recruited for this study: 7 right-handed English monolingual speakers (5 females, mean age=25.0 years (sd=5.1)); 7 right-handed native Portuguese speakers (4 Portuguese monolinguals and 3 Portuguese-English bilingual speakers, 3 females, mean age=25.1 years (sd=3.8)); and 7 right-handed native Mandarin speakers (all Mandarin-English bilinguals, 5 females, mean age=23.9 years (sd=2.6)). A previous study (Yang et al., 2017) showed that the neural representations of bilinguals and monolinguals reading sentences in their native language have comparable decoding accuracies. All participants signed informed consent approved by the Carnegie Mellon University Institutional Review Board (IRB protocol HS14-474). They reported normal or corrected-to-normal vision and no history of traumatic head injuries. All the participants read the stimulus sentences in their native languages.

2.2 Stimulus materials

Sixty English sentences, part of a larger set from a study with English monolingual speakers (Glasgow, Roos, Haufler, Chevillet, & Wolmetz, 2016; Wang et al., 2017) were translated into Portuguese and Mandarin (e.g. *The mayor negotiated with the mob/O prefeito negociou com a multidão/市长和暴徒谈判*) by native speakers. These sentence stimuli obey the SVO (subject-verb-object) order and have a mean length of 3.2 content words. The original English sentences were constructed systematically to establish a stimulus set for evaluating models of neural decoding of sentences (Glasgow, Roos, Haufler,

Table 1. Language combinations in *Two-to-One Mappings*, *One-to-One Mappings,* and *Within-Language Mappings*.

	Training Language	**Test Language**
Two-to-One Mappings	English & Portuguese	Mandarin
	English & Mandarin	Portuguese
	Portuguese & Mandarin	English
One-to-One Mappings	English	Portuguese
	Portuguese	English
	English	Mandarin
	Mandarin	English
	Portuguese	Mandarin
	Mandarin	Portuguese
Within-Language Mappings	English	English
	Portuguese	Portuguese
	Mandarin	Mandarin

Chevillet, & Wolmetz, 2016). A complete set of the translation equivalents of the 60 sentences appears in Table S.1 of the Supplemental Materials.

An independent group of Portuguese-English and Mandarin-English bilinguals (none of whom were participants in the fMRI testing) agreed on the translation equivalencies of the sentences in the two languages. To further confirm the translation equivalence of the sentences, the Portuguese and Mandarin sentences were translated back to English by an independent group of bilinguals, and the resulting English sentences were judged for their semantic equivalency to the original English sentences by an English monolingual speaker. This process was repeated until the back-translated English sentences were all judged to be semantically equivalent to the original English sentences.

2.3 Experimental Paradigm

The participants read the 60 sentences in their native language while fMRI scans were acquired. The sentences were presented in a white font against a black background. Each sentence was divided into phrases at natural phrase boundary in the respective language (e.g. The family, was, happy/ *A família, estava, feliz/这家人,挺,快乐*). Each phrase was presented one at a time, left justified, in a moving window format, with a speed that is comparable to the natural reading speed of native speakers in the respective language (Just and Carpenter 1980, 1987). The adjectives in noun phrases remained on the screen together with the nouns they modified, until the nouns disappeared. Participants were instructed to read each phrase silently and think about the meaning, and integrate it to the conception of the whole sentence as it unfolded. At the end of the sentence, a blank interval padded out the total presentation duration to 5 sec. A schematic depiction of the presentation of one sentence in each of the three languages is shown in Figure 1.

During the blank interval at the end of each sentence, participants were instructed to continue thinking about the sentence, integrating the meaning of all the words. After each blank interval, an X appeared at the center of the screen for 7 s during which participants were instructed to fixate and clear their minds. Each one hour fMRI session included four presentation blocks of the 60 sentences presented in different random orders. There were also 16 occurrences of 17-second fixation periods, distributed across the entire session, to provide a baseline measure of activation.

2.4 fMRI acquisition and processing

Functional images were acquired on a Siemens Verio 3.0T scanner at the Scientific Imaging & Brain Research Center (SIBR) of Carnegie Mellon University (gradient echo EPI pulse sequence; TR = 1000 ms, TE = 30 ms, and a 60º flip angle). Twenty 5-mm thick AC-PC aligned slices were imaged (1-mm gap between slices). The acquisition matrix was 64 x 64 with 3.125 x 3.125 x 5-mm voxels.

The data were corrected for head motion and normalized to the Montreal Neurological Institute (MNI) template using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). For each presentation of a sentence, the percent signal change (PSC) was computed at each voxel, relative to the mean baseline activation level of the voxel measured during fixation periods. The MPSC (mean PSC) for each voxel for each sentence in each presentation was computed as the mean of five PSC images, collected from 7 seconds to 11 seconds post sentence onset (TR=1 sec). This temporal window was found to be the most decodable window as determined by a previous study (Wang et al. 2017). The MPSC images were then normalized to a mean of 0 and variance of 1 across voxels for each sentence. Because each sentence in each language was presented 4 times, four MPSC images of each sentence of each language were obtained.

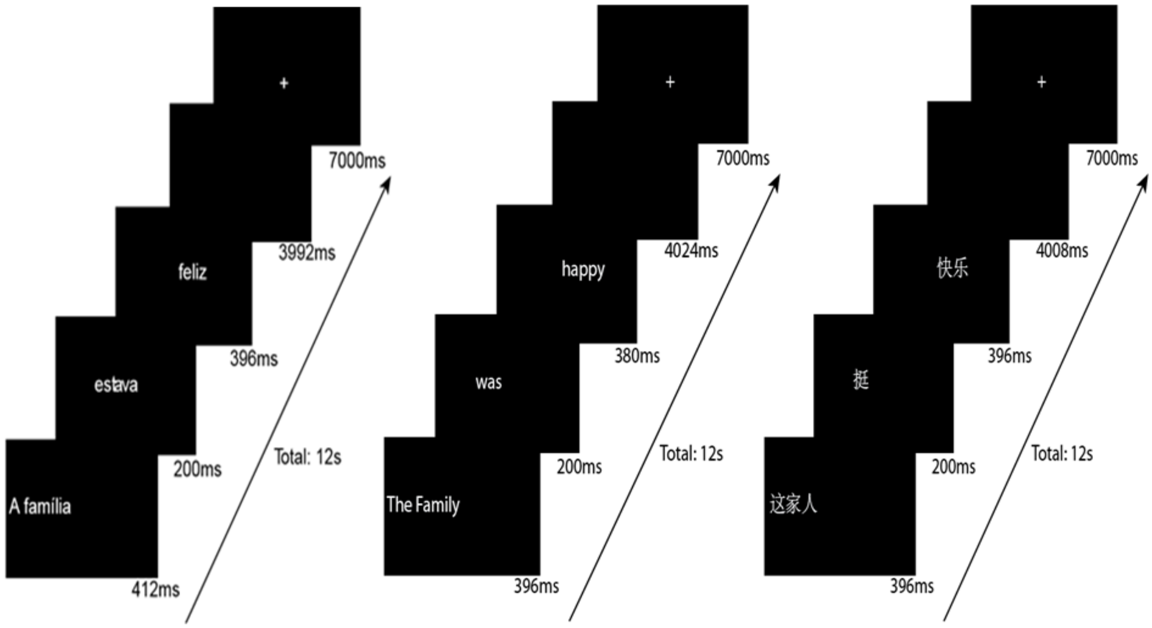

Fig. 1. The presentation paradigm of one sample sentence in Portuguese, English and Mandarin. The presentation duration of each phrase was determined by a regression model based on previous eye movement studies of text reading. The duration was specified as 300 ms × number of words+ 16 ms× number of letters in each content word for Portuguese and English, and 300 ms× number of words + 8 ms ×number of strokes in each content character for Mandarin

2.5 Defining activation regions common across languages

First, within each language, the voxels that had the most reliable tuning curves across the 60 sentences were identified by computing stability scores for each voxel in each participant. Occipital lobe voxels were excluded from the analysis to decrease the possible effects of low-level visual features of the printed sentences. The stability score of a voxel was defined as the average of the six pairwise correlations of the voxel's 60 MPSC activation levels (for the 60 sentences) between the six possible pairings of two of the four presentations. Then the distributions of stability scores across different languages and participants were examined to select a threshold such that roughly similar numbers of voxels in the 250 range were above this threshold in all three languages. Then, voxels that had stability scores higher than this threshold (0.08 on a normalized MPSC image) were identified in each participant. Finally, the stable voxels that were common (in MNI space) across individual participants of each language group were identified, resulting in 246 voxel locations in the Portuguese dataset, 260 in the English and 268 in the Mandarin. These shared stable voxels were clustered by the SPM function *spm_clusters*, to obtain language-specific clusters, as shown in Figure 2A. A set of language-general clusters was formed by combining the three language-specific clusters, as shown in Figure 2B. (The three averaged and clustered language-specific stability maps were averaged and thresholded to obtain a roughly similar total numbers of voxels (250) as there were in each of the three languages.)

The irregularly-shaped language-general clusters were grown into the smallest cuboids that fully enclosed each cluster, to regularize the volumes under consideration. The language-general clusters (black voxels) and the boundaries of the grown cuboids (red contours) are shown on a glass brain in Figure 2C.

2.6 Gaussian Naïve Bayes Classifications of Two-to-One Mapping and One-to-One Mapping

A discriminative Gaussian Naïve Bayesian (GNB) classifier was trained on the mapping between the neural activation patterns of the 60 sentences and the sentence labels. The training data were obtained from either a single language (e.g. Portuguese) or two languages (e.g. English and Portuguese). The classifier was then tested on the neural activation patterns of the sentences obtained in another language (e.g. Mandarin). We will refer to training on activation patterns in a single language to classify activation patterns in another as a *One-to-One Mapping* and training on activation patterns from two languages to classify activation patterns in a third language as a *Two-to-One Mapping*. For the three languages examined here, there are six possible *One-to-One Mappings* and three possible *Two-to-One Mappings,* as shown in Table 1. In addition, a within-language between-participant classification was conducted in each language (using the same amount of training and test data as in the other mappings), to provide an estimate of the upper bound of classification accuracy that cross-language classifications can be expected to reach.

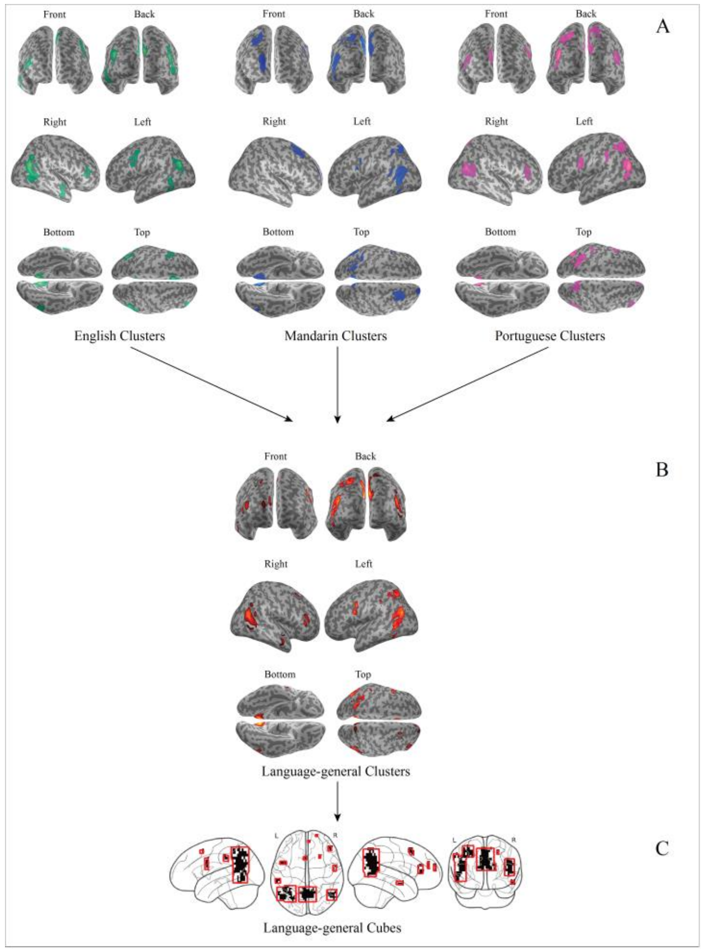

Fig. 2. (A) Language-specific clusters of stable voxels that were common across participants. English clusters in green (left), Mandarin clusters in blue (middle), and Portuguese clusters in pink (right); (B) Language-general clusters formed by combining the language specific clusters; (C) The smallest cuboids (red outlines) that contain each language-general cluster, shown on a glass brain.

The number of training participants and training scans were equated between *Two-to-One, One-to-One* and within-language classifications to make the classification accuracies comparable. To equate the training participants, three participants from each one of the two training languages were selected in each fold of the *Two-to-One Mappings* (sampling all possible combinations of three participants from the two languages across folds), and six participants from the training language were selected for *One-to-One Mappings* and within-language mappings for each *training iteration*. Each training iteration contained six nested *classification folds*. In each *classification fold*, two out of the four scans of the 60 sentences from each training participant were used to train the classifier and the mean of two scans of the test participant was used to test classifier. The scans of the same sentence were assigned the same sentence label, resulting in 60 sentence labels in each language. The three types of classifications all attempted to identify each of the 60 test sentences from its MPSC activation pattern in the cuboids described above.

The classifier's features that characterized each sentence were the mean activation levels of the representative voxels in the 15 language-general cuboids. The representative voxels in each classification fold were the 15% of the voxels in each language-general cuboid with the highest stability. The activation levels of these selected voxels within a cuboid were averaged to obtain the activation level characterizing the cuboid for each sentence of each individual participant. These activation levels were then concatenated as a vector (one vector for each sentence) and served as the features to train the classifier. In the training phase, two scans of each sentence per training participant were used. Thus, there were 12 example activation vectors (2 scans times 6 participants) on which the classifier was trained to learn the activation pattern evoked by each sentence.

The test data for the classifier for each sentence was the mean of the activation patterns in the language-general clusters of two test scans of the test subject from the test language. The classifier placed each test sentence into a rank ordered list of the 60 sentence labels, ordered by the classifier's probability ranking. Classification accuracy for each sentence was computed as the normalized rank of the target sentence among all the 60 sentence candidates.

This normalized rank is defined as rank accuracy and was implemented as:

$$\text{Rank accuracy} = \frac{(total\ number\ of\ sentences\ -\ the\ rank\ of\ the\ correct\ target)}{(total\ number\ of\ sentences\ -1)}$$

For each participant, the rank accuracies were averaged across all sentences in all classification folds and all training iterations. To compare the classification accuracies across different conditions (e.g. *Two-to-One* vs. *One-to-One* mappings), ANOVAs were used.

2.7 The semantic domain analyses of the two-language advantage

To determine whether any accuracy difference between *Two-to-One* vs. *One-to-One* mappings was related to specific types of semantic properties, the rank accuracies for each sentence in a given test language were compared between the two types of mappings (e.g. Portuguese and English to classify Mandarin vs Portuguese only to classify Mandarin). This resulted in 60 sentence-wise accuracy differences in 6 contrast pairs (e.g. training on Portuguese and English to classify Mandarin minus training on Portuguese to classify Mandarin, training on Portuguese and English to classify Mandarin minus training on English to classify Mandarin, and so on). The 6 contrast pairs were averaged to obtain the overall sentence accuracy differences between *Two-to-One* and *One-to-One* mappings.

These 60 sentence accuracy differences were then regressed against a set of word-level Neurally Plausible Semantic Features (NPSFs) to determine which NPSFs benefitted most from the advantage of the *Two-to-One Mapping* over the *One-to-One Mapping.* (The sentence-level NPSFs were computed by summing up the NPSFs of all component words of the sentence). The NPSFs are both interpretable (e.g. *shelter*, *tools, eating/drinking*, and *emotion*) and have a neural signature (as described in Yang et al., 2016). They were coded as binary semantic features at a word level and designed to characterize semantic concepts that are neurally plausible across languages (e.g. *doctor* was coded as *Animate* and *Person* in all three languages; *computer* was coded as *Manmade* and *Technology* in all three languages, etc. (See Table S.2 for a list of the NPSFs and the coding of some sample concepts). The NPSFs corresponding to each concept in the sentences were coded by two English native speakers in English, and were normed and checked by native speakers of Mandarin and Portuguese to make sure that the coding was correct for each of these languages. For each sentence, the NPSF codes of all the individual content words in each sentence were added and treated as the independent variables that characterize the set of concepts in a sentence. A similar set of neurally plausible features has been proposed by other researchers (Binder et al., 2016), and these features were also demonstrated to be useful in sentence classification in English (Anderson et al., 2016).

2.8 Representational Similarity Analysis of the Sentence Semantics

The neural activation patterns displayed by the voxels used in the classification were also used to conduct a Representational Similarity Analysis (RSA) of the sentence semantics (Haxby et al. 2001; Nili et al. 2014; Guntupalli et al. 2016) within each of the three languages (English, Portuguese, Mandarin). Inter-sentence similarities were computed as the cosine similarity between the neural activation patterns of the selected voxels from each participant. Then the individual RSA matrices were averaged into a mean RSA matrix for each language. This resulted in three 60 x 60 RSA symmetrical matrices for the three languages (symmetrical because the similarity between sentence i and j is the same as the similarity between sentence j and i, and the entries on the diagonal are between identical sentences). To determine if the set of inter-sentence neural similarities was similar across pairs of languages, the set of inter-sentence similarities within each language was correlated with each of the other two languages (removing the redundant and self-similar entries).

3. Results

3.1 Advantage of the Two Languages-to-One Mapping over One-to-One Mapping

The *Two languages-to-One* mappings (training on two languages and testing on the third language) produced a significantly higher mean classification accuracy (.668) than the *One-to-One* mappings (training on one language and testing on another) (.624) ($F(1,61)=19.4$, $p<0.001$). These means were obtained by averaging over all possible combinations of training and test languages, so the mean for the *Two languages-to-One* mappings, for example, is averaged over three cases, where either English, Portuguese, or Mandarin is the test language and the other two are the training languages. The mean sentence classification accuracies for each such case (averaged across test participants, training iterations and classification folds) are summarized in Table 2. It is notable that the two-language advantage is apparent in the test of all three languages.

All the cross-language and within-language classification accuracies (regardless of the number of training languages) were significantly above chance level. (The critical value for $p = 0.05$ level accuracy is 0.56, estimated by a 5000-iteration random permutation test). The mean classification accuracies were slightly higher when English was one of the two training languages or the only training language than in the other cases (.65 vs .63, $F(1,61)=7.77$, $p < .05$)

The mean within-language cross-participant classification accuracy of .669 using the same amount of training data indicates the upper bound that cross-language classification should be able to reach. The *Two-to-One Mappings* reached this upper bound (i.e. the *Two-to-One* classification accuracies are almost equal to the within-language classification accuracies) in all cases while the *One-to-One Mapping* classification accuracies were significantly lower than this ceiling ($F(1,61)=20.63$, $p < 0.001$).

There was a small but reliable classification accuracy advantage when English was a training language, attributable to all the English participants having had prior experience in performing similar semantic tasks in fMRI studies, facilitating their adjustment to the fMRI environment and their engagement with the task. One measure of the English speakers' expertise as fMRI participants was their lower amount of head motion (mean overall displacement) than the other two groups, reliably so with respect to the Portuguese participants ($t(12)=2.87$, $p<0.014$) and marginally so with respect to the Mandarin participants ($t(12)=2.06$, $p<0.062$). The lower amount of head motion would have resulted in less noise in the data of the English participants and hence greater classification accuracy.

Table 2. Classification accuracy of *Two-to-One, One-to-One, and Within-Language* mappings. (Standard deviations across test participants indicated in parentheses).

	Training Language(s)	Test Language	Classification Accuracy	Mean
Two-to-One Mappings	English & Portuguese	Mandarin	0.67 (0.03)	.668
	English & Mandarin	Portuguese	0.67 (0.05)	
	Portuguese & Mandarin	English	0.66 (0.04)	
One-to-One Mappings	English	Portuguese	0.63 (0.05)	.624
	Portuguese	English	0.63 (0.02)	
	English	Mandarin	0.65 (0.03)	
	Mandarin	English	0.63 (0.03)	
	Portuguese	Mandarin	0.61 (0.04)	
	Mandarin	Portuguese	0.60 (0.03)	
Within-Language Mappings	English	English	0.66 (0.04)	.669
	Portuguese	Portuguese	0.67 (0.06)	
	Mandarin	Mandarin	0.67 (0.08)	

3.2 Relating the two-language advantage to semantic features

The regression of the sentence-wise accuracy differences between the *Two-to-One* and *One-to-One* mappings against NPSFs showed that 5 NPSFs reliably benefitted from the classifier having been trained on two languages. The NPSFs that reliably benefitted (at $p < 0.05$, significance level determined by a permutation test of randomizing the NPSF labels 1000 times) from the *Two-to-One* mapping are ***Abstraction*** (as coded in words like *wealthy, happy, like, negotiation*), ***Person*** (as coded in words like *mob, artist, banker, author*), ***Communication*** (as coded in words like *listen, negotiation, shout),* ***Social*** (as coded in words like *couple, soccer, famous),* and ***Knowledge*** (coded in words like *author, engineer, negotiate*). The NPSFs that did *not* benefit from *Two-to-One* mapping include ***Man-made*** (coded in words like *glass*, *desk*, *dime, car*), ***Setting*** (coded in words like *school*, *theater*, *field*), ***Natural*** (coded in words like *dog*, *horse*), and ***Visual Perception*** (coded in words like *big*, *shiny*), etc. This outcome indicates that the decoding of abstract and social concepts benefits most from training a classifier on two languages, and suggests that it is these types of concepts that are somewhat differently neurally represented in different languages.

3.3 Assessing the neural similarity of inter-sentence relations across languages

To compare the concept representation spaces across languages, Representational Similarity Analyses (RSA) were first conducted on the sentences within each language (as shown in Supplementary Figures S.1, S.2, and S.3), and then the inter-sentence similarity measures were compared across languages. Two large clusters of sentences emerged in all three languages, together involving 35 of the 60 sentences. One large cluster contains 16 sentences describing *Event in an Environmental Scene,* shown as a red region of highly correlated sentences in the upper left of each array in Figure 3. The other large cluster contains 19 sentences *Social Interaction,* shown as a red region of highly correlated sentences in the lower right of each array in Figure 3. The emergence of the same two sentence clusters in the three panels of Figure 3 illustrates that the inter-language neural similarities for all three pairs of languages are similar to each other.

The correlations between the three pairs of 60 x 60 RSA matrices were all reliable ($p < 0.001$), showing a great amount of cross-language similarity. These correlations were: English-Portuguese .63; English-Mandarin .60; Mandarin-Portuguese .59. The correlation between the English and Portuguese matrices was slightly higher than the correlation between Portuguese and Mandarin matrices ($p=0.04$), indicating that the neural semantic space is more similar between Portuguese and English than between Portuguese and Mandarin. While the cross-language classifications illustrate the first-order similarity between the neural representations of sentences across languages, the RSA's illustrate the second-order similarity of the inter-sentence neural similarities across languages.

4. Discussion

4.1 Main implications

The main contribution of the current study is its comparison of pairwise neural commonality across languages versus commonality across three languages. Previous cross-language decoding studies (Buchweitz et al., 2012; Zinszer et al., 2012, 2016; Yang et al., 2016) have indicated that neural concept representation and processing have a great deal of commonality

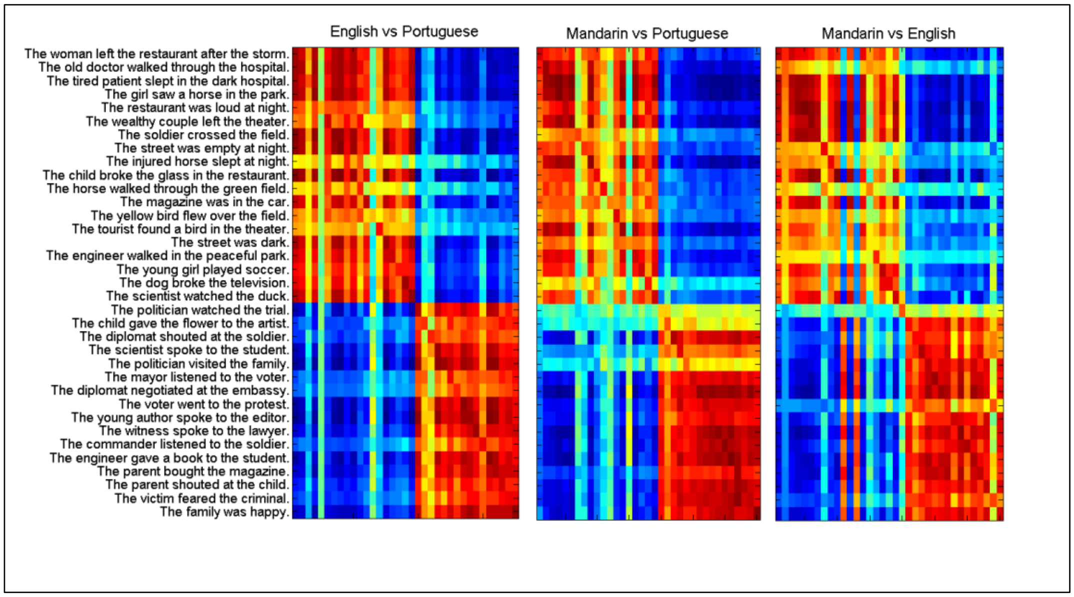

Fig. 3. Two common cross-language clusters emerged from the matrices of the pairwise correlations between the RSA matrices in each language. Left panel: correlation matrix between English RSA and Portuguese RSA matrices; middle panel: correlation matrix between Mandarin RSA and Portuguese RSA matrices; right panel: correlation matrix between Mandarin RSA and English RSA matrices. The top left submatrix constitutes the cluster of sentences describing events in an environmental scene; the bottom right submatrix constitutes the cluster of sentences describing social interactions.

across pairs of different languages: similar brain locations are activated in the processing of similar concepts, and the patterns of neural activation levels evoked by equivalent concepts within these common brain locations are similar. This study extended this meta-language commonality to three languages, affording the opportunity to assess the specificity of pairwise neural commonalities across languages. The results clearly show that there is reliably more to the cross-language neural commonality than what a pairwise comparison reveals.

The main finding was that the classification accuracy was reliably higher when the classifier was trained on two languages compared to having been trained on one language, even when the amount of training data was equated. Furthermore, the accuracy of the classifier trained on two languages matched the accuracy level of the within-language classification (which defines an upper bound on cross-language classification accuracy). It is possible that if the stimulus sentences contained more culture-specific information than those in the current study, even a Two-to-One mapping would fail to capture all of the within-language mapping between meaning and activation pattern, while still producing a higher accuracy than a One-to-One mapping.

The advantage of the classifier trained on two languages stems from the inclusion of the neural mappings that are common only to its second training language and the test language, thus enlarging the training domain, as illustrated by the region shown in black in Figure 4. The results of the study show that language-specific overlap mappings exist and that they differ from language to language. There is reliably more to the cross-language neural commonality than can be found with a pairwise comparison between languages.

4.2 What types of information does a second training language contribute?

The finding of a two-language advantage raises the question of what types of concepts compose the regions of pair-wise *only* overlap between languages (illustrated by the region shown in black in Figure 4). A second new finding of this study was that the sentences whose accuracy benefitted most from the classifier having been trained on two languages were those that contained concepts that are more abstract and were related to social and mental activities e.g. *happy*, *negotiation*, *artist* (coded with NPSFs like ***Abstract***, ***Communication***, ***Mental action or State***).

The reason for this differential benefit may be that abstract and socially-related concept domains (particularly of the type of abstract concepts sampled by the stimulus sentences) have a greater degree of cultural or language-specific influence that is reflected in the neural representations. Training a classifier on two languages includes neural mappings of such concepts in two languages, increasing the probability that one of them will provide a good match to the neural representations in the test language. Thus the region of overlap between only the second training language and the test language, shown in black in Figure 4, may contain disproportionately many abstract concepts. (Of course, if the comparison was between the accuracy of training a classifier on both languages versus only on training language 2, then the pairwise-only additional overlap would be the region with diagonal stripes, and that region, like the black one would presumably also be composed of disproportionately-many sentences containing abstract concepts).

Although abstract concepts may have made a large contribution to the two-language advantage, it is difficult to specify the content of an abstract concept, which probably varies by type of abstract concept. Abstract concepts are often defined by what they are not, namely they do not have a concrete, perceptible referent. At the core of an abstract concept is a

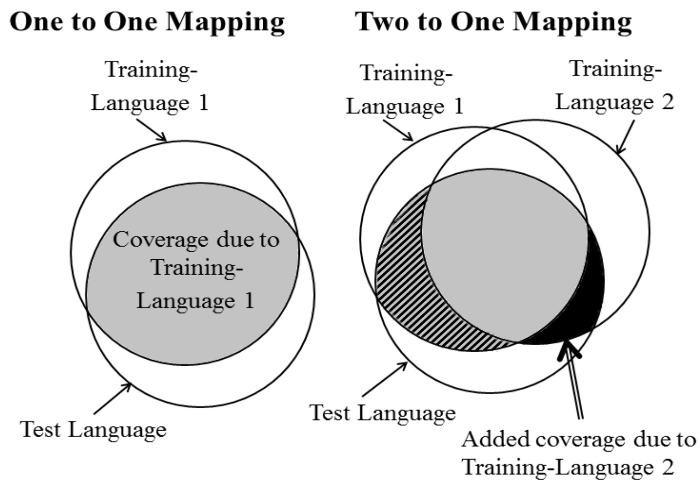

Fig. 4. Illustration of the concept-to-neural mapping domains among the three languages. The additional information provided by a Two-to-One mapping compared to One-to-One mapping is shown in black.

definition in terms of other concepts. In this sense, an abstract concept is one that is composed of the meanings of other concepts. Some abstract concepts pertain to human traits and social interactions. Some abstract concepts pertain to affect and emotion. Some abstract concepts are part of a formalized system, such as abstract scientific concepts. Although the stimulus materials do not lend themselves to a definitive conclusion about the content of the abstract concept representations in this study that underpinned the two-language advantage, a few tentative inferences can be made.

The results do not support the hypothesis that affective components of the neural representations played a substantial role in the advantage of training a classifier on two languages. The affective components of some words may have been language specific (for example, *dog,* could evoke more positive affect in cultures in which dogs are domesticated and negative affect in cultures where dogs are scavengers). However, none of the NPSF's related to affect showed a reliable two-language advantage.

The stimulus sentences contained few scientific concepts, which are typically abstract but nevertheless precise and presumably common across cultures. It is possible that the neural representations of sentences describing scientific concepts would *not* show a benefit of a classifier being trained on two languages, because all of the relevant information would be contained in the neural representations of the concepts in any single training language (all other things being equal). The critical variable may not be whether a concept is abstract but whether there is room for variation in its meaning across languages.

Although we have been assuming that the mappings between sentences and their neural representations are based primarily on the semantic content of the sentences, it is possible to consider whether lower levels of information about the sentence (such as its articulatory or syntactic properties) may also have been involved in the mapping between neural representations and sentences in a way that contributes to classification accuracy. The situation is different for the 12 different mappings shown in Table 2, because the relevance of such lower level information depends on the similarity between languages in such lower level features. The most relevant cases arise in the variation among the six One-to-One mappings. The similarity between English and Portuguese in such features is greater than between Mandarin and English or between Mandarin and Portuguese. The classification accuracies in Table 2 for the One-to-One mappings bear no relation to the similarities between languages. The mean accuracy for the four One-to-One mappings involving Mandarin (.62) is very similar to the mean accuracy for two One-to-One mappings between English and Portuguese (.63). Thus it seems unlikely that low level features contributed to the mapping where it was possible for them to have done so, namely between English and Portuguese. Given that lower level features did not contribute to the accuracy in the One-to-One mappings, it seems very unlikely that contributed to the Two-to-One mappings. If they happened to play a role in the within-language mappings, that would only have made it more difficult for the Two-to-One mappings to achieve comparable accuracies, which they nevertheless did.

4.3 Inter-sentence neural similarity

The similarity in the neural activation patterns across languages is further indicated by RSA analyses, which showed significant correlations between within-language inter-sentence similarities across all three pairs of languages. The between-language correlations of inter-sentence similarities produced two large clusters of sentences. The emergence of common clusters in all pairs of languages that described *events in environmental scenes* or that described *social interactions* indicates that these two types of sentences evoke distinct neural representations that are similarly related to each other in all three languages.

The findings above implying that abstract and social concepts are neurally represented somewhat differently across languages is consistent with the RSA findings of similar clustering of sentences describing social interactions across languages (coded by NPSFs like ***Communication***, ***Person***, and ***Knowledge***). Even though the sentences describing social interactions were clustered in the RSA, their mean degree of between-language correlation (.61) was reliably lower than for the *events in environmental scenes* cluster (.82), ($p < 0.001$). This outcome is in accordance with the above two-language advantage results: the neural representations of sentences describing social interactions are more language specific than sentences describing events in environmental scenes.

4.4 Limitations of the study

Comparisons of neural representations of concepts across languages are always influenced by the choice of participants. Some participants in this study were bilingual (all Mandarin participants and 3 Portuguese participants) and the remainder were monolingual. Several previous studies found no differences between bilingual and monolingual speakers in their neural representations of concepts they were reading in their L1 (Buchweitz and Prat, 2013; Correia et al., 2014; Kovelman et al., 2007; Yang et al., 2016). Specifically, Yang et al. (2016) demonstrated almost identical within-language classification accuracies for 8 Portuguese-English bilinguals and 7 Portuguese monolingual participants when they were reading Portuguese sentences. This finding indicates that being bilingual seems not to affect the neural representations evoked by L1 in this paradigm.

A second caveat in any cross-language comparison concerns the incomplete equivalence of translation-equivalent words and sentences (Panou 2013). Minor lexical differences in the translation equivalent sentences may also have contributed to the two-language advantage. For example, there is no exact equivalent of *dime* in Portuguese, which was translated as ten cents. Another possibility is that the neural representation for the identical referent could be different in two languages if it were viewed differently in different cultures, such as the word *pork* evoking a different response in cultures in which it is a forbidden food. Because languages are never identical, their neural representations cannot be identical, and the challenge becomes to identify the sources of systematic differences.

Third, it is possible in principle that the small amount of variation in the stimulus sentence lengths could have affected the neural representations, and the classifier could have used any such correlation to contribute to accuracy. However, the classification analyses excluded occipital lobe voxels to decrease the possible effects of visual properties features of the printed sentences, such as their lengths. To eliminate the possibility that sentence length played a role, the classification within English was repeated, but with sentence length regressed out (i.e. the classification was done on the residuals of the sentence length regression analysis). The resulting classification accuracy was almost identical (.661 vs. .662) to the analysis without the correction for sentence length. Given the absence of a length effect on classification accuracy within a language, where it would be expected to be largest, this additional result indicates that sentence length information did not contribute to classification accuracy in any of the analyses and particularly not to the Two-to-One advantage.

Finally, the sample size of participants, stimulus materials, and languages in the current study is relatively small. A larger sample in all three respects would further illuminate the cross-language commonality issue. It would be particularly interesting to systematically vary the types of sentence structures and content to assess the impact on the cross-language commonalities, using our approach.

4.5 Conclusion

The human brain provides a common neural platform for representing sentences in all languages, resulting in a great deal of commonality in such representations across languages. At the same time, each language and culture can introduce nuances in the meaning and hence neural representation of concepts that superficially seem similar. Examining the mappings between concepts and neural representations in multiple languages has the potential for revealing the existence of such language specificities and the semantic domains in which they tend to occur. Identifying both the universals as well as the language specificities is necessary for characterizing the full range of mappings between brain and language.

Acknowledgements

This work was supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Air Force Research Laboratory (AFRL) contract number FA8650-13-C-7360. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. During data collection, Cyntia Bailer was supported by the Brazilian Ministry of Education, CAPES BEX 14636-13-1. We thank Nick Diana, Robert Vargas, and Zachary Anderson for their help in data collection, stimulus preparation and coding, and Leda Tomitch for guidance on bilingual issues.
Corresponding author: Marcel Adam Just, Carnegie Mellon University, Department of Psychology, Pittsburgh, PA 15213;
E-mail: just@cmu.edu.

References

Athanasopoulos, P., & Kasai, C. (2008). Language and thought in bilinguals: The case of grammatical number and nonverbal classification preferences. *Applied Psycholinguistics*, *29*(1), 105–123.

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., ... & Raizada, R. D. (2016). Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cerebral Cortex*.

Berman, R. A., & Slobin, D. I. (2013). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3-4), 130-174.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22. http://doi.org/10.1006/cogp.2001.0748

Buchweitz, A., & Prat, C. (2013). The bilingual brain: Flexibility and control in the human cortex. *Physics of Life Reviews*, *10*(4), 428–443. http://doi.org/10.1016/j.plrev.2013.07.020

Buchweitz, A., Shinkareva, S. V., Mason, R. a., Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language*, *120*(3), 282–289. http://doi.org/10.1016/j.bandl.2011.09.003

Cook, V., Bassetti, B., Kasai, C., Sasaki, M., & Takahashi, J. (2006). Do bilinguals have different concepts? The case of shape and material in Japanese L2 users of English. *International Journal of Bilingualism*, *10*(2), 137–152. http://doi.org/10.1177/13670069060100020201

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 : The Official Journal of the Society for Neuroscience*, *34*(1), 332–8. http://doi.org/10.1523/JNEUROSCI.1302-13.2014

Enfield, N. J. (2015). *Natural causes of language: Frames, biases, and cultural transmission* (p. 97). Language Science Press.

Frankland, S.M., & Greene, J. D., (2014). An architecture for encoding sentence meaning in left mid-superior temporal cortex, PNAS, 112(37), 11732-11737.

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A Model of Representational Spaces in Human Cortex. *Cerebral Cortex*, bhw068. http://doi.org/10.1093/cercor/bhw068

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(September), 2425–2430. http://doi.org/10.1126/science.1063736

Han, S., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience*, *9*(8), 646-654.

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.

Johansson, A., & Salminen, S. (1996). Different languages: different information processing systems? Part I: Why can we expect differences in occupational accidents between language groups. In *Int. Symp. Work Inf. Soc., Helsinki*.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*. US: American Psychological Association. http://doi.org/10.1037/0033-295X.87.4.329

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Allyn & Bacon.

Kovelman, I., Baker, S. A., & Petitto, L.-A. (2007). Bilingual and Monolingual Brains Compared: A Functional Magnetic Resonance Imaging Investigation of Syntactic Processing and a Possible "Neural Signature" of Bilingualism. *Journal of Cognitive Neuroscience*, *20*(1), 153–169. http://doi.org/10.1162/jocn.2008.20011

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition*, *84*(2), 155–188. http://doi.org/10.1016/S0010-0277(02)00045-8

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1-20.

Lucy, J. A. (2016). Linguistic Relativity Author (s): John A . Lucy Published by : Annual Reviews Stable URL : http://www.jstor.org/stable/2952524 REFERENCES Linked references are available on JSTOR for this article : LINGUISTIC RELATIVITY, *26*(May), 291–312.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4). http://doi.org/10.1371/journal.pcbi.1003553

Ozgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: a perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology. General*, *131*(4), 477–493. http://doi.org/10.1037/0096-3445.131.4.477

Panou, D. (2013). Equivalence in translation theories: A critical evaluation. *Theory and Practice in Language Studies*, *3*(1), 1.

Strømnes, F. J. (1974). To be is not always to be The hypothesis of cognitive universality in the light of studies on elliptic language behaviour. *Scandinavian Journal of Psychology*, *15*(1), 89–98.

Wang, Q., & Conway, M. A. (2004). The stories we keep: Autobiographical memory in American and Chinese middle-aged adults. *Journal of Personality*, *72*(5), 911–938. http://doi.org/10.1111/j.0022-3506.2004.00285.x

Wang, J., Cherkassky, V. L., & Just, M. A. (2017). Predicting the brain activation pattern associated with the propositional content of a sentence: Modeling neural representations of events and states. *Human Brain Mapping.*

Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. oxford university Press.

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *NeuroImage*, *146*, 658-666.

Zinszer, B. D., Anderson, A. J., Kang, O., & Wheatley, T. (2012). How speakers of different languages share the same concept, 2829–2834

Zinszer, B. D., Anderson, A. J., & Raizada, R. D. (2016) Chinese and English speakers' neural representations of word meaning offer a different picture of cross-language semantics than corpus and behavioral measures. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Supplemental Materials

Table S.1. The 60 translation equivalent stimulus sentences in English, Portuguese and Mandarin

ID	English	Portuguese	Mandarin
1	The family was happy.	A família estava feliz.	这家人挺快乐
2	The politician visited the family.	O político visitou a família.	政客拜访了这家人
3	The family played at the beach.	A família brincou na praia.	这家人在海滩玩
4	The parents bought the magazine.	Os pais compraram a revista.	父母买了那本杂志
5	The child broke the glass in the restaurant.	A criança quebrou o copo no restaurante.	孩子在餐馆打碎了玻璃杯
6	The parents shouted at the child.	Os pais gritaram com a criança.	父母对着孩子嚷
7	The happy couple visited the embassy.	O casal feliz visitou a embaixada.	那对幸福的夫妻拜访了大使馆
8	The wealthy couple left the theater.	O casal rico saiu do teatro.	那对有钱的夫妻离开了剧院
9	The parents visited the school.	Os pais visitaram a escola.	父母参观了学校
10	The happy child found the dime.	A criança feliz encontrou a moeda de dez centavos.	那个快乐的小孩发现了一枚十分硬币
11	The child gave the flower to the artist.	A criança deu a flor para o artista.	孩子给了艺术家一朵花
12	The soldier crossed the field.	O soldado atravessou o campo.	士兵穿越了田野
13	The commander listened to the soldier.	O comandante ouviu o soldado.	指挥官聆听士兵
14	The horse walked through the green field.	O cavalo atravessou o campo verde.	那匹马穿过了那片绿色的田野
15	The girl saw a horse in the park.	A garota viu um cavalo no parque.	女孩在公园里看见一匹马
16	The engineer walked in the peaceful park.	O engenheiro caminhou no parque tranquilo.	工程师在平静的公园里走着
17	The flower was yellow.	A flor era amarela.	花是黄色的
18	The yellow bird flew over the field.	O pássaro amarelo sobrevoou o campo.	黄色的鸟飞越田野
19	The old doctor walked through the hospital.	O velho médico andou pelo hospital.	年老的医生穿过医院
20	The wealthy author walked into the office.	O autor rico entrou no escritório.	富有的作家走进了办公室
21	The dog broke the television.	O cachorro quebrou a televisão.	狗打坏了电视
22	The street was empty at night.	A rua estava vazia à noite.	街道晚上空无一人
23	The street was dark.	A rua estava escura.	街道很昏暗
24	The banker watched the peaceful protest.	O banqueiro assistiu ao protesto pacífico.	银行家观看了和平抗议活动
25	The voter went to the protest.	O eleitor foi ao protesto.	那个选民去了抗议
26	The protest was loud.	O protesto foi barulhento.	抗议声音很大
27	The politician watched the trial.	O político assistiu ao julgamento.	政治家观看了审判
28	The reporter spoke to the loud mob.	O repórter falou com a multidão barulhenta.	记者和吵闹的暴民谈话
29	The mayor negotiated with the mob.	O prefeito negociou com a multidão.	市长和暴民谈判
30	The mob was dangerous.	A multidão era perigosa.	暴民很危险
31	The wealthy politician liked coffee.	O político rico gostava de café.	那个富有的政治家很喜欢咖啡
32	The young author spoke to the editor.	O autor jovem falou com o editor.	年轻的作家和编辑谈话
33	The scientist spoke to the student.	O cientista falou com o estudante.	科学家和学生谈话
34	The scientist watched the duck.	O cientista observou o pato.	科学家观察鸭子
35	The witness went to the trial.	A testemunha foi ao julgamento.	证人去了审判
36	The witness spoke to the lawyer.	A testemunha falou com o advogado.	证人和律师谈话

37	The witness shouted during the trial.	A testemunha gritou durante o julgamento.	证人在审判中叫嚷
38	The jury watched the witness.	O júri observou a testemunha.	陪审团观察证人
39	The victim feared the criminal.	A vítima temia o criminoso.	被害人恐惧罪犯
40	The engineer gave a book to the student.	O engenheiro deu um livro para o estudante.	工程师给了学生一本书
41	The magazine was in the car.	A revista estava no carro.	杂志在车里
42	The diplomat negotiated at the embassy.	O diplomata negociou na embaixada.	外交官在使馆交涉
43	The diplomat shouted at the soldier.	O diplomata gritou com o soldado.	外交官对士兵嚷
44	The mayor listened to the voter.	O prefeito ouviu o eleitor.	市长聆听选民
45	The famous diplomat left the hospital.	O diplomata famoso deixou o hospital.	那个著名的外交官离开了医院
46	The patient survived.	O paciente sobreviveu.	病人活下来了
47	The tired patient slept in the dark hospital.	O paciente cansado dormiu no hospital escuro.	疲惫的病人在昏暗的医院里睡
48	The author kicked the desk.	O autor chutou a mesa.	作家踢了书桌
49	The tourist went to the restaurant.	O turista foi ao restaurante.	游客去了餐馆
50	The woman left the restaurant after the storm.	A mulher saiu do restaurante depois da tempestade.	那个女人在暴风雨后离开了餐馆
51	The restaurant was loud at night.	O restaurante estava barulhento à noite.	餐馆晚上很热闹
52	The artist liked chicken.	O artista gostava de frango.	艺术家喜欢鸡肉
53	The diplomat was wealthy.	O diplomata era rico.	外交官很富有
54	The mayor dropped the glass.	O prefeito derrubou o copo.	市长掉了玻璃杯
55	The injured horse slept at night.	O cavalo machucado dormiu à noite.	受伤的马晚上睡了
56	The young girl played soccer.	A jovem garota jogou futebol.	这个年轻的女孩踢足球
57	The girl saw the small bird.	A garota viu o passarinho.	女孩看见那只小鸟
58	The tourist found a bird in the theater.	O turista encontrou um pássaro no teatro.	游客在剧院发现一只鸟
59	The school was famous.	A escola era famosa.	这学校很有名
60	The magazine was yellow.	A revista era amarela.	那本杂志是黄颜色的

Table S.2. Definitions of Neurally Plausible Semantic Features (NPSF) and examples of words coded

Category *	Feature	Definition	Examples
Perceptual and Affective Characteristics of an Entity	Man-made	objects or settings made by humans	*bicycle, desk, newspaper, church*
	Natural	objects or activities occurring in nature	*flower, flood, island*
	Inanimate	non-living object	*ball, coffee, window*
	Visual perception	visual perceptual properties	*big, blue, empty, new, shiny*
	Size	Physical volume or size	*big, heavy, long, small*
	Color	self-explanatory	*black, blue, green, red, white*
	Temperature	related to temperature	*sun, summer, winter, cold, hot*
	Positive valence	self-explanatory	*celebrate, laugh, vacation, happy*
	Negative valence	self-explanatory	*destroy, fear, terrorist, dangerous, sick*
	High intensity	high affective arousal	*celebrate, shout, hurricane, angry*
Animate Beings	Person	a human being	*boy, doctor, farmer, pilot, voter*
	Animal	an animal or anatomy of animals	*bird, dog, duck, feather, horse*
	Human-group	more than one human being	*team, couple, family, mob, council*
Time and Space	Settings	place or temporal settings	*lake, church, park, night, hotel*
	Unenclosed	an environment without shelter or enclosure	*beach, lake, field, island, street*
	Location	actions or events that imply spatial settings	*meeting, visit, stay, live*
	Shelter	being enclosed, indoors is a salient feature; opposite of unenclosed	*car, hotel, school, hospital, store*
	Change of location	self-explanatory	*approach, hike, throw, car, run*
	Event	self-explanatory	*dinner, protest, trial, vacation*
	Time-related	related to a time period or timing	*morning, night, spring, summer, end*
Human Activity Type	Violence/conflict	involving aggression and those who commit it	*army, guard, soldier, terrorist*
	Health	related to improving or threatening health	*medicine, doctor, patient, victim, hospital*
	Eating/drinking	self-explanatory	*drink, eat, dinner, corn, restaurant*
	Communication	medium of communication	*listen, speak, newspaper, author, reporter*
	Sports	related to recreation or competitive physical activities	*play, soccer, baseball, bicycle, team*
	Technology	related to technology or technical skills	*computer, television, engineer, scientist*
	Money	related to financial activities or economics	*buy, cash, banker, expensive, wealthy*
	Arts and literature	objects, actions or professions related to humanities, arts, literature	*actor, author, artist, theatre, draw*
	Social norms	related to law and authority structure	*trial, criminal, lawyer, court, prison*
	Governance	related to civics, politics, dominance	*debate, protest, army, mayor, embassy*
	Intellectual	requiring, gaining, or providing knowledge or expertise	*plan, read, computer, engineer, school*
Social Action or State	Transfer of possession	transaction (giving/receiving); change of ownership	*give, steal, take, buy*
	Social interaction	interaction between two or more subjects	*interview, negotiate, party, lonely*
	Social support	providing social support is a salient feature	*help, family, minister, parent*
Physical Action or State	Physical action	self-explanatory	*kick, throw, play, walk, march*
	Change of physical state	self-explanatory	*destroy, fix, grow, break*
	Physical impact	two subjects or objects coming in contact with each other	*break, destroy, drop, kick*

Mental Action or State	Mental action	requiring cognitive processes; occurring internally	*liked, plan, want, teacher, clever*
	Perceptual action	self-explanatory	*listen, watch, read, witness*
	Emotion	Emotional state or action	*fear, laugh, like, happy*
Abstractness	Abstract	detached from sensory or motor properties; low imageability	*plan, want, clever*
Part of Speech	Attribute	adjectives	*aggressive, blue, shiny, sick*

* The grouping into categories is intended to facilitate description and is not used in the modeling.

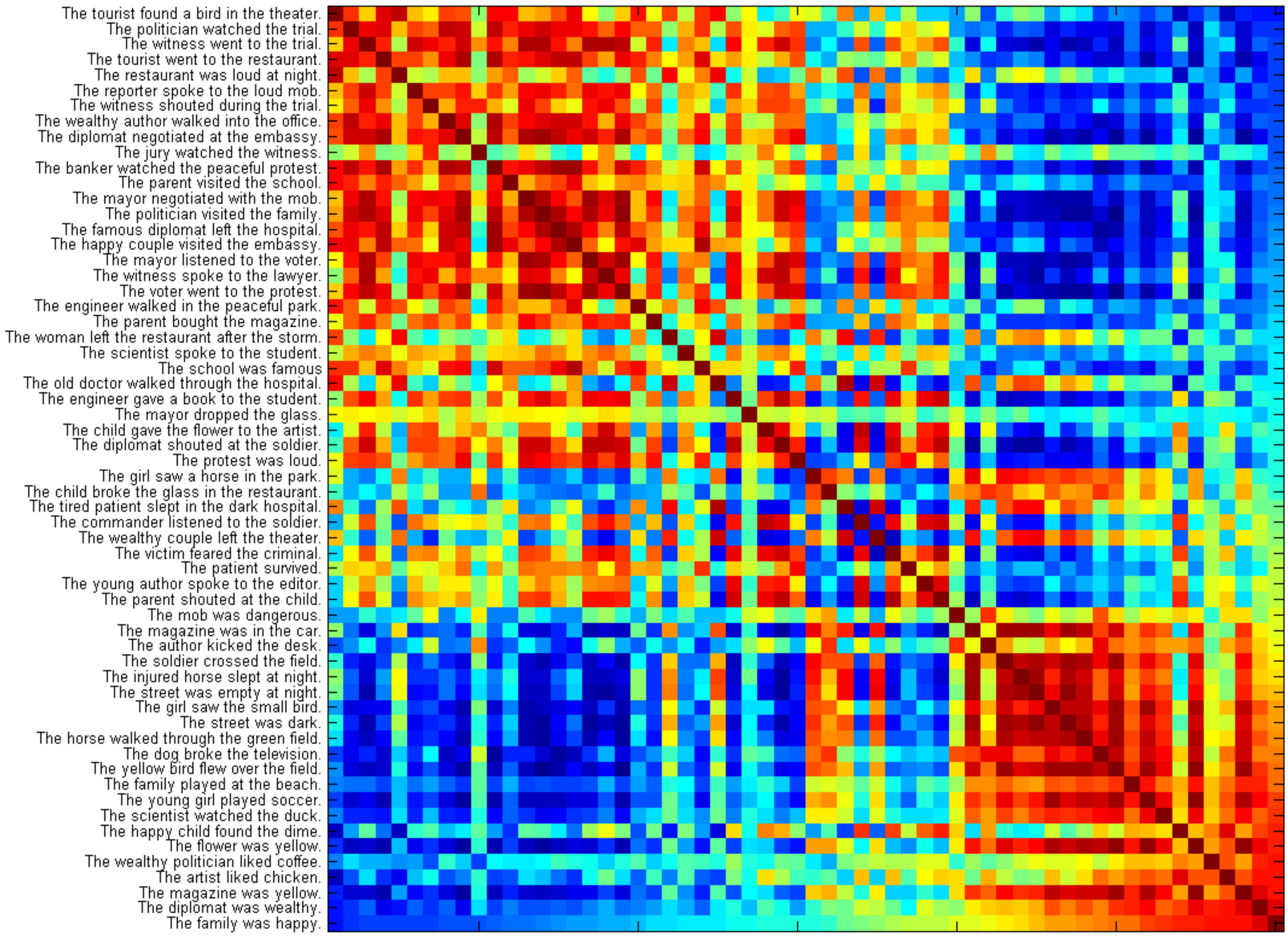

Figure S.1. Representational Similarity Matrix of neural representations of English sentences whose similarities formed clusters, sorted by the row similarities. Two clusters emerged: one describing *Event in an environmental scene* (in the upper left corner), and the other describing *Social Interactions* (in the lower right corner)

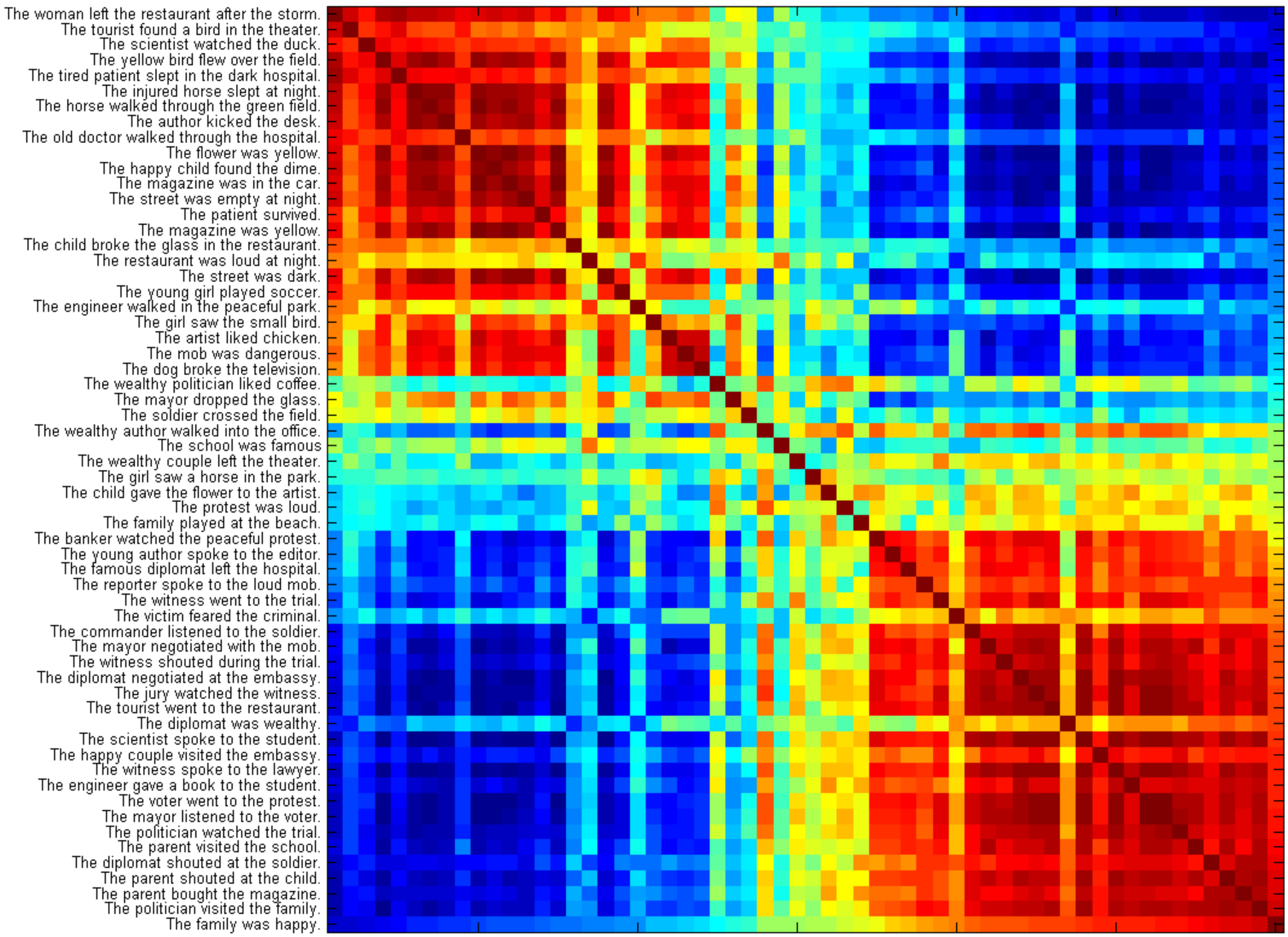

Figure S.2. Representational Similarity Matrix of neural representations of Portuguese sentences whose similarities formed clusters, sorted by the row similarities. Two clusters emerged: one describing *Event in an environmental scene* (in the upper left corner), and the other describing *Social Interactions* (in the lower right corner)

Figure S.3. Representational Similarity Matrix of neural representations of Mandarin sentences whose similarities formed clusters, sorted by the row similarities. Two clusters emerged: one describing *Event in an environmental scene* (in the upper left corner), and the other describing *Social Interactions* (in the lower right corner)