Brain & Language 215 (2021) 104922

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Brain decoding in multiple languages: Can cross-language brain decoding work?

Min Xu[a,b,*], Duo Li[c], Ping Li[c,*]

[a] Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen 518060, China
[b] Center for Language and Brain, Shenzhen Institute of Neuroscience, Shenzhen 518060, China
[c] Department of Chinese and Bilingual Studies, Faculty of Humanities, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

ARTICLE INFO

Keywords:
Cross-language brain decoding
Neural representation
Multivariate pattern analysis
Computational modeling
Multilingualism

ABSTRACT

The approach of cross-language brain decoding is to use models of brain decoding from one language to decode stimuli of another language. It has the potential to provide new insights into how our brain represents multiple languages. While it is possible to decode semantic information across different languages from neuroimaging data, the approach's overall success remains to be tested and depends on a number of factors such as cross-language similarity, age of acquisition/proficiency levels, and depth of language processing. We expect to see continued progress in this domain, from a traditional focus on words and concrete concepts toward the use of naturalistic experimental tasks involving higher-level language processing (e.g., discourse processing). The approach can also be applied to understand how cross-modal, cross-cultural, and other nonlinguistic factors may influence neural representations of different languages. This article provides an overview of cross-language brain decoding with suggestions for future research directions.

1. Introduction

Decoding language from neural activity has become an exciting and challenging research topic, in large part due to the rapid advances in artificial intelligence, and in brain-inspired computing, that is, using what is known about the brain for the design of novel computing systems (Anumanchipalli, Chartier, & Chang, 2019; Poldrack, 2011). 'Brain decoding of language', as a relatively new field of research, refers to the following approach of study (see Fig. 1 for illustration): neural responses to linguistic materials are recorded with neuroimaging methods, such as functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG); a computational model is trained to map between brain activity and stimulus-specific linguistic features; if the model successfully predicts new linguistic stimuli from brain activity, it means that the model captures important semantic-conceptual features of the stimuli, thereby achieving the purpose of decoding the language stimuli. The dimensions needed to capture the semantic features of the stimuli (i.e., the semantic space) may be postulated by researchers or derived from text corpora which can be based on statistical regularities inherent in the text (e.g., word co-occurrences). Many factors determine the accuracy of brain decoding, including the temporal and spatial resolutions in the neuroimaging recordings and the type and nature of the computational model (Gallant, 2016). The approach of brain decoding of language not only helps us to understand how the brain represents language, but also has important clinical and educational implications. For example, it could be used to predict what words a person is hearing, reading or even thinking, which, in the future, could inform the design of brain-computer interfaces. The proper development of such interfaces can make a huge difference in people's lives, especially for those who suffer from communication disorders, including aphasia due to stroke or other neurodegenerative diseases.

A landmark study of brain decoding of language was conducted by Mitchell et al. (2008), which showed that it was possible to accurately predict which concrete concepts (e.g., celery) a participant was thinking of by analyzing the corresponding brain activations in response to the concepts (e.g., a picture of celery). The predictions were made based on a linear regression model that was trained to establish the mapping between the intermediate semantic features of the input nouns and the corresponding brain activation elicited by the nouns. The intermediate semantic features that were used to encodes the meaning of the nouns were defined by 25 verbs, such as *eat, taste, smell, hear, see, touch* and so on, and the feature values (the co-occurrence of the semantic features

* Corresponding authors at: Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong 518060, China (M. Xu). Faculty of Humanities (GH801), The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China (P. Li).

E-mail addresses: xumin@szu.edu.cn (M. Xu), pi2li@polyu.edu.hk (P. Li).

https://doi.org/10.1016/j.bandl.2021.104922
Received 18 July 2020; Received in revised form 5 January 2021; Accepted 19 January 2021
Available online 5 February 2021

and the stimulus words) were obtained from a very large corpus of text. The brain activations were recorded with fMRI while participants watched repeated images presented in the scanner. The model, with above-chance accuracy, not only reliably predicted the brain activation elicited by novel nouns but also decoded the nouns from a new dataset based on brain activation.

An exciting new direction in recent years has been cross-language brain decoding, which is our focus here. This direction of research helps us to reveal how our brain represents multiple languages. Traditional neuroimaging studies of bilingualism have compared neural activities elicited by different languages, and identified both common and distinct neural systems of multiple languages (see Li, 2009, for a discussion). The cross-language brain decoding approach provides a new and powerful direction to address the issue of how two or more languages are encoded through shared and distinct neural activities in the brain. These studies have significant practical implications for bilingual education and foreign language instruction. For example, the study of different ages of acquisition or proficiency levels of second language learners would allow us to disentangle whether and how much the linguistic background and experience might influence the success of cross-language decoding in the bilingual brain, which could potentially inform us about mechanisms underlying critical periods for language acquisition. However, so far it is unclear whether and how cross-language brain decoding works, given the extant evidence. Simply put, can we use models of brain decoding from language A and apply them successfully to decode language B, and vice versa? In this article, we provide a review of recent studies of brain decoding in different languages in an attempt to identify the various factors that may affect the success of cross-language decoding. We will focus on cross-language decoding at the word level, discuss sentence- and passage-level decoding, and conclude with a discussion of future directions.

2. Cross-language brain decoding at the word level

2.1. Approaches and findings

A number of recent studies have demonstrated that it is possible to reliably decode semantic information at the word level across different languages from neuroimaging data using machine learning methods. The general approach is to first train a decoder to learn the mappings between stimuli in language A (e.g., English) and the corresponding brain activity elicited by the stimuli. Then, if the decoder is successful for language A, it is tested on a new set of data, this time the brain activity evoked by stimuli from a new language, language B (e.g., Chinese) (Fig. 2). Multivariate pattern analysis (MVPA) has been used in cross-language decoding with increasing popularity (Haxby, 2012). Compared to the traditional univariate method which examines brain voxels in isolation, MVPA takes into account the relationships across multiple voxels and has the potential to decode fine-grained patterns of brain activity. Table 1 presents a summary of the cross-language brain decoding studies that vary in the use of participants, materials, tasks, and data analytic methods (most of which had used MVPA). Variations along each of these dimensions could affect the predictive accuracy in the studies. In what follows, we take a detailed examination of these studies.

Bilinguals are usually recruited as participants in cross-language brain decoding studies and the same participants need to receive stimuli (words) from both languages (consecutively) while their brain responses are collected during the processing of these stimuli. Buchweitz, Shinkareva, Mason, Mitchell, and Just (2012) used concrete nouns from two categories (tools and dwellings) as stimuli and presented the nouns in Portuguese and English (translation equivalents) consecutively to Portuguese-English bilinguals. Participants were required to read each noun silently and think about the properties of the noun while their brain activity was recorded using fMRI. Results showed that, when the decoder was trained on the fMRI signals elicited by the English nouns and tested on the fMRI activity elicited by the Portuguese nouns, the decoding accuracy reached 0.68. Likewise, when the decoder was

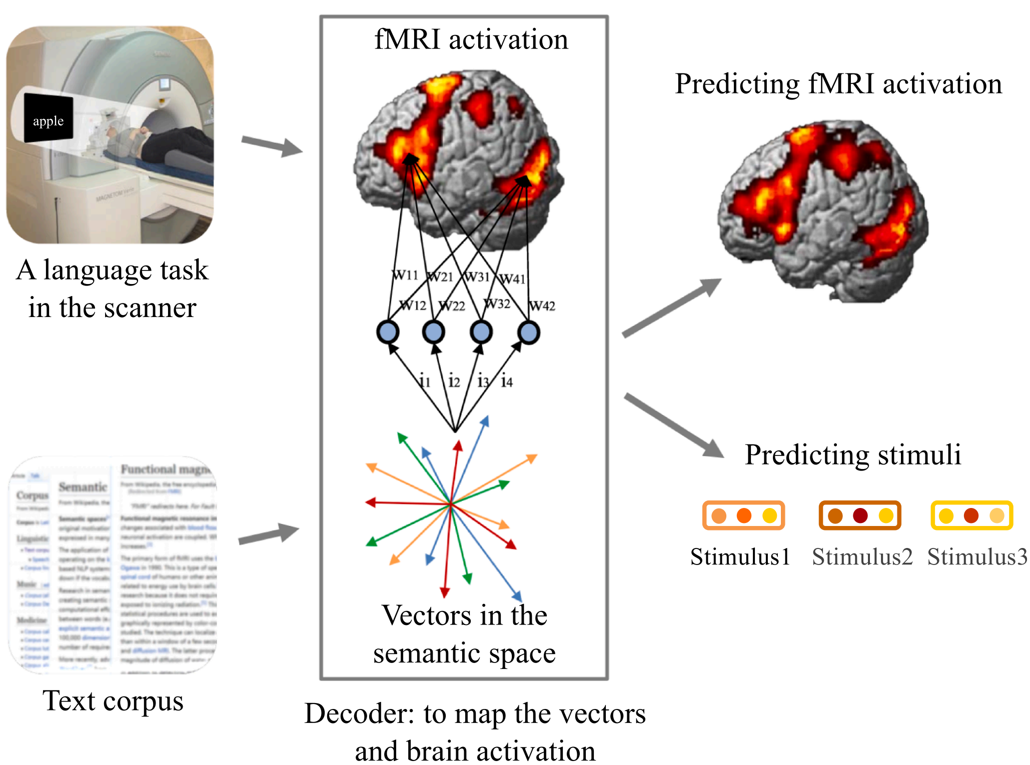

Fig. 1. General approach for brain decoding of language. Brain activation of linguistic stimuli can be recorded when participants perform a language task in the scanner. Vector-based, high-dimensional representations of the linguistic stimuli can be derived from a large text corpus. A decoder is trained to map the semantic vectors of the stimuli and the corresponding brain activation. The decoder can then be applied to predict what brain activation may result given a stimulus, or predict which of the stimulus it is given the specific brain activation patterns.

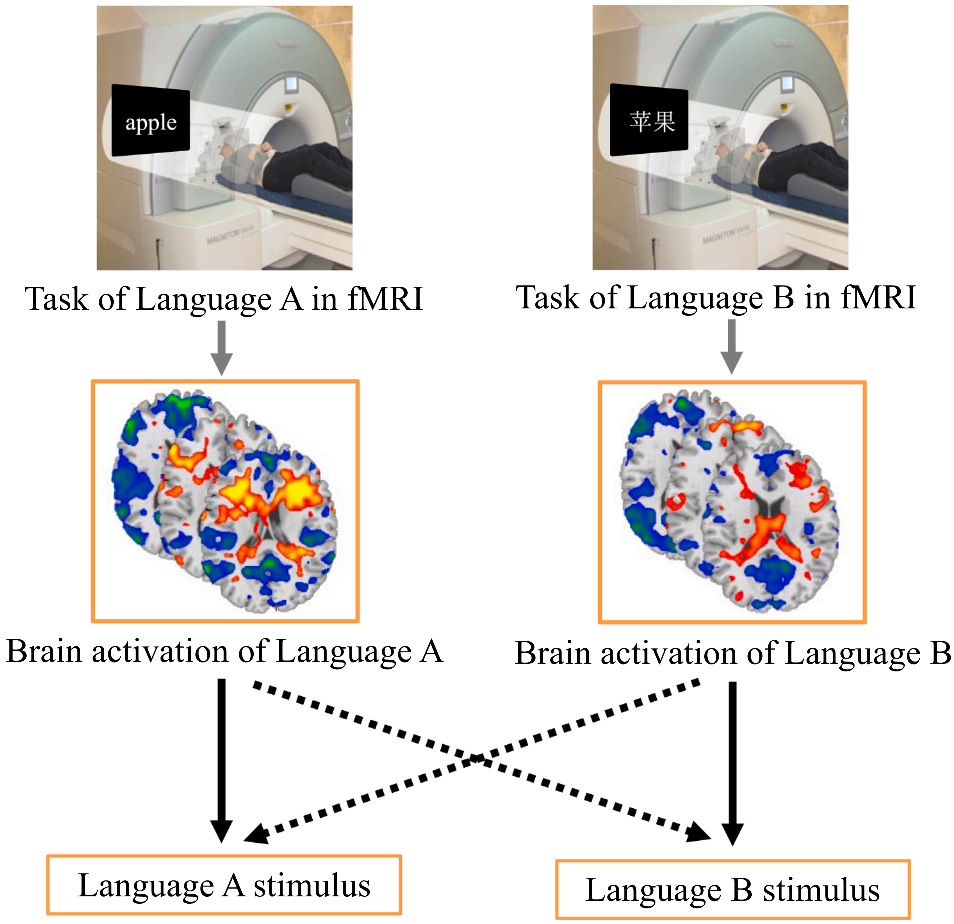

Fig. 2. An illustration for within- and cross-language brain decoding. Solid black arrows indicate within-language decoding, and dotted black arrows indicate cross-language decoding.

trained with the Portuguese nouns and tested on the brain activity elicited by the English nouns, the decoding accuracy reached 0.72. Both were significantly higher than the chance level (i.e., 0.5). The authors suggested that the representation of the meanings of the same nouns may share the same neural substrates between English and Portuguese.

Subsequent research further confirmed the feasibility of cross-language brain decoding at the word level. Correia et al. (2014) examined the decoding of spoken words between Dutch and English. Dutch-English bilinguals were asked to listen to animate and inanimate nouns and press a button when they heard the inanimate nouns in the MRI scanner. The detection accuracy was 97.5%, indicating that participants knew the words in both languages. MVPA revealed several regions responsible for the cross-language word discrimination, including the anterior temporal lobe. Their follow-up study (Correia, Jansma, Hausfeld, Kikkert, & Bonte, 2015) measured EEG signals to identify the time course of Dutch-English cross-language decoding. Although within-language decoding might have taken place around 50–620 ms after the word onset, cross-language decoding occurred later, around 550–600 ms after the word onset. According to the authors, this broad timeline for within-language decoding and narrow time window for cross-language decoding could be explained by the different mechanisms underlying decoding: within-language decoding relies on the initial phonetic-phonological processing and the subsequent lexical semantic processing, whereas cross-language decoding could rely on the shared semantic-conceptual properties of the word across languages, therefore occurring within a narrow but fast time window.

Cross-language decoding success in bilinguals reported in the above studies reflects the common neural representation of word meanings across two languages. However, since the participants recruited in these studies were proficient bilinguals who know the words in both languages very well, it is likely that cross-language decoding reflects the association of translation equivalents in bilinguals rather than true similarities in the neural substrates of representation. This hypothesis is especially plausible given the Correia et al. (2015) finding that the cross-language decoding time window is significantly shorter than the within-language decoding time window. In contrast to the use of bilinguals, Zinszer, Anderson, Kang, Wheatley, and Raizada (2015) and Zinszer, Anderson, Kang, Wheatley, and Raizada (2016) tested only monolinguals so that the Chinese-speaking participants read only Chinese words (e.g., 'fu' 斧) and English-speaking participants read only English words (e.g., 'axe'). They were asked to determine whether the word was semantically related to the preceding word in the MRI scanner. Neural similarity matrices were calculated for each participant based on activation patterns in their own language. Surprisingly, neural similarity matrices for English and Chinese were strongly correlated ($r = 0.89$, $p < 0.001$), thus allowing accurate cross-language decoding. For some brain regions the decoding accuracy reached 100% (Zinszer et al., 2015), including the anterior parahippocampal and postcentral gyrus in the left hemisphere and the frontal orbital cortex, anterior cingulate gyrus, anterior supramarginal gyrus and posterior inferior temporal gyrus in the right hemisphere. However, this study tested only seven Chinese-English words (translation equivalents) referring to concrete objects, and therefore its generalizability remains unclear.

2.2. Applications to other domains

The cross-language brain decoding approach has not only been applied to data from language comprehension (in which participants

Table 1
Summary of studies of cross-language brain decoding.

Study	Participant	Mean Age	AoA	Proficiency in L2	Method	Stimuli and task	Informative clusters	Cross-language decoding accuracy	Within-language decoding accuracy
Buchweitz et al., 2012	11 Portuguese-English BI	29.9	13.08	Proficient (Self-rated)	fMRI; MVPA	Concrete nouns; Thinking about the properties of nouns	Left postcentral and supramarginal gyri, inferior/superior parietal lobes, inferior frontal gyrus and posterior superior temporal lobe	English → Portuguese 0.68, Portuguese → English 0.72 (CL 0.5)	English 0.60, Portuguese 0.63 (CL 0.5)
Correia et al., 2014	10 Dutch-English BI	25.4	–	Proficient (LexTALE*)	fMRI; MVPA	Concrete nouns; Pressing a button when they hear an inanimate noun	Left anterior temporal lobe	Above chance (CL 0.5)	Above chance (CL 0.5)
Correia et al., 2015	16 Dutch-English BI	28.9	–	Proficient (LexTALE)	EEG; MVPA	Concrete nouns; Pressing a button when hear an inanimate noun	–	Above chance during 550–600 ms after word onset (CL 0.5)	Above chance during 50–620 ms after word onset (CL 0.5)
Zinszer et al., 2015	11 English native speakers and 11 Chinese-English BI	–	–	–	fMRI; MVPA	Concrete nouns; Determining whether the noun was semantically related to the preceding word	Left anterior parahippocampal and postcentral gyrus, right frontal orbital cortex, anterior cingulate, supramarginal and inferior temporal gyri	The accuracy reached 1.0 for some ROIs (CL 0.5)	–
Van de Putte et al., 2017	24 Dutch-French BI	23.4	0 for early BI, 9 for late BI	Different levels of proficiency (LexTALE, BNT** and self-rating)	fMRI; MVPA	Pictures; Naming the pictures	Bilateral occipito-temporal cortex, inferior/middle temporal gyri	0.110 (CL 0.1)	–
Van de Putte et al., 2018	22 Dutch-French BI	23.6	0 for early BI, 9 for late BI	Different levels of proficiency (LexTALE, BNT and self-rating)	fMRI; MVPA	Pictures and concrete nouns; Naming the pictures and determining the properties of the concepts	Occipito-temporal cortex, rolandic operculum, pre- and postcentral, cerebellum	Above chance (CL 0.1)	–
Sheikh et al., 2019a	30 Spanish-Basque BI	24.2	0.24 for Spanish, 1.17 for Basque	Proficient, more proficient in Spanish (LexTALE, BEST***)	fMRI; MVPA	Words; Shallow processing task: reading and attending to the word; Deep processing task: thinking about the properties of the word	Left inferior parietal lobe, lateral and ventromedial temporal lobe, inferior frontal and posterior cingulate gyri	Deep processing: Spanish → Basque 0.550, Basque → Spanish 0.548, shallow processing: chance-level (CL 0.5)	Deep processing: Spanish 0.597, Basque 0.600, shallow processing: Spanish 0.582, Basque 0.576 (CL 0.5)
Sheikh et al., 2019b	24 Spanish-Basque BI	22.3	0.52 for Spanish, 1.05 for Basque	Proficient; more proficient in Spanish (LexTALE, BEST)	fMRI; MVPA	Words; Determining whether the word was animate or inanimate and rating the awareness of the word	Left inferior parietal lobe, dorsomedial prefrontal cortex, inferior frontal, posterior cingulate and ventromedial temporal cortices	Non-conscious trials: CL, partially conscious trials: CL (CL 0.5)	Non-conscious: Spanish 0.546, Basque 0.546, partially conscious: Spanish 0.539, Basque 0.537 (CL 0.5)
Yang et al., 2017a	8 Portuguese-English BI and 7 Portuguese MO	28.1	12.9	Highly proficient (adapted TOEFL)	fMRI; semantic features	Sentences; Thinking about the properties of each phrase	Multiple brain regions	English → Portuguese 0.67 (CL 0.5)	–
Yang et al., 2017b	7 English MO, 4 Portuguese MO, 3 Portuguese-English BI and 7 Mandarin-English BI	24.7	–	–	fMRI; semantic features	Sentences; thinking about the sentences	Multiple brain regions	Two-to-one mappings 0.668 on average, one-to-one mappings 0.624 on average (CL 0.5)	English 0.66, Portuguese 0.67, Chinese 0.67 (CL 0.5)
Dehghani et al., 2017	30 English MO, 30 Chinese-English BI and 30 Farsi-English BI	24.7	–	Fluent	fMRI; story-level embeddings (doc2vec)	Narrative stories; Reading the stories	Default mode network (posterior medial cortices, medial prefrontal and lateral parietal cortices)	One-to-one mappings 0.561 on average (CL 0.5)	English 0.566, Farsi 0.549, Chinese 0.552 (CL 0.5)

Note: →indicates the direction of cross-language brain decoding (e.g., English → Portuguese indicates the models for brain decoding English is applied to decode Portuguese). Abbreviations: AoA: age of acquisition; BI: bilinguals; MO: monolinguals; CL: chance level.
*LexTALE is a vocabulary knowledge test (Lemhöfer & Broersma, 2012).
**BNT (Boston Naming test) is a picture naming test that measures word retrieval ability (Kaplan, Goodglass, & Weintraub, 2001).
***BEST (Basque, English, and Spanish tests) is a series of tests of language proficiency (De Bruin, Carreiras, & Duñabeitia, 2017).

listened or read the stimulus words) as reported above, but also from language production (in which participants named the stimuli). For example, Van de Putte, De Baene, Brass, and Duyck (2017) asked Dutch-French bilinguals to name pictures of 10 concepts in Dutch and French consecutively. The order of presenting the two languages was counterbalanced across participants. Two different images were selected to represent each concept and each image was associated with one language in order to exclude the influence of visual similarity. For example, a moon was represented by a crescent moon in the Dutch naming blocks (e.g., 10 concepts to be named in Dutch) and a full moon in the French naming blocks. Their data indicated above-chance level accuracy for cross-language decoding in the left inferior and middle temporal gyrus and the bilateral occipitotemporal cortex, suggesting shared semantic representations of L1 and L2 word production in these regions. However, this study didn't find the involvement of frontal regions and anterior temporal regions, in contrast to the role of these regions for cross-language decoding in studies that have used comprehension tasks (Buchweitz et al., 2012; Correia et al., 2014). Such discrepancies may suggest different neural representations associated with production and comprehension, which affect accuracy of cross-language decoding.

In the same spirit as brain decoding across languages, brain decoding has also been applied to different modalities. For example, Van de Putte, De Baene, Price, and Duyck (2018) investigated the decoding across visual and auditory modalities, in which Dutch-French bilinguals were instructed to complete three different tasks in Dutch and French consecutively. In essence, this is both cross-modality and cross-language decoding. The picture naming task was the same as that used in the above-mentioned study of Van de Putte et al. (2017), and the word reading task (visual) and word-listening task (auditory) required participants to make a judgement about the properties of the concept (e.g., whether the concept was bigger or smaller than a football). Significant cross-language predictions were observed in the rolandic operculum and some motor-related areas (pre- and post-central, the cerebellum) in both word reading and listening tasks. More interestingly, it was possible to identify the picture the participant was naming (auditory) in one language based on the brain activation elicited by the reading of the corresponding word (visual) in the other language, and vice versa. The cross-modality decoding effect was most pronounced in the left lingual gyrus, suggesting its critical role in language-independent semantic processing. This result suggested the existence of both modality-independent and modality-dependent semantic representation, but the specific brain regions that support language-independent semantic processing may vary. According to the 'hub-and-spoke' model (Ralph, Jefferies, Patterson, & Rogers, 2017), modality-specific and modality-independent representations are realized in different neural circuits, in visual/auditory/motor areas versus anterior temporal lobe, respectively. The left lingual gyrus could be part of the modality-independent neural network responsible for visual and motoric information processing.

Another potentially exciting domain of application of cross-language decoding is sign language (SL) vs. spoken language decoding, although there has been limited work so far on this topic. Despite the clear differences in input/output modes, the neural substrates of SL and spoken language appear to both involve a predominantly left-lateralized brain network as revealed by lesion and neuroimaging studies (Emmorey, Giezen, & Gollan, 2016; Macsweeney, Capek, Campbell, & Woll, 2008). By examining cross-language decoding between SL and spoken language, we can determine (a) the extent to which the neural representations of language are dependent on (or independent of) the modality, and (b) if and how differences can arise due to modality-specific representations. Comparison between this new line of research with extant work from unimodal bilinguals could offer a unique opportunity to understand the nature of cross-modality conceptual representation.

3. What factors affect cross-language decoding success?

The above review suggests that cross-language brain decoding is a promising approach to understand how multiple languages are encoded in the human brain. However, the decoding accuracy varies across studies, and is affected by many factors. In what follows, we provide a synthesis of the potential factors that may underlie the degree of success in this approach.

3.1. Cross-language similarity

One factor that affects cross-language decoding is the distance or similarity between the two languages, which may involve systematic differences in vocabulary, grammar, phonology, script, and other characteristics. For instance, English and Chinese are distant in terms of both spoken and written forms (Li, Tan, Bates, & Tzeng, 2006). Chinese is a tonal language in which different pitches and duration of the sound convey different meanings of words, and is a non-inflectional language that relies heavily on contextual semantics. By contrast, English is a non-tonal language, and uses inflectional morphemes to assign grammatical properties to words (e.g., tense morphology).

Understanding the extent to which there are shared or different aspects across languages is an essential step in addressing the possible influences of language properties and linguistic experience on cross-language brain decoding. One well-studied example of crosslinguistic differences would be linguistic categories, including lexical categories and grammatical categories (e.g., Malt & Majid, 2013; Pavlenko, 2009). There are many cases in which the structure or boundaries of lexical categories do not neatly match across languages, even for concrete nouns referring to common objects whose appearances and functions are expected to be generally the same in different languages. For instance, clear cross-language differences have been shown in the naming of common household items (such as cups, dishes, bottles and jars) in speakers of English, Chinese, and Spanish (Malt, Sloman, Gennari, Shi, & Wang, 1999; Malt, Sloman, & Gennari, 2003) and French and Dutch (Ameel, Storms, Malt, & Sloman, 2005). Such differences also have direct influences on bilingual speakers' naming performance, as shown in both behavioral studies and computational models (Fang, Zinszer, Malt, & Li, 2016; Malt, Jobe, Li, Pavlenko, & Ameel, 2016; Malt, Li, Pavlenko, Zhu, & Ameel, 2015). Besides, in the classic case of color terms, languages can differ in the size of color vocabularies in that some languages have a relatively larger set of color items that divide the color spectrum more finely than other languages. The English category blue, for example, is lexically differentiated by Russian and Greek speakers using different terms to describe dark blue and light blue (Athanasopoulos, 2009; Pavlenko, 2009).

With respect to grammatical categories like nouns and verbs, the distinction between noun and verb classes is transparent in some languages (such as German), but it could be ambiguous in other languages because of the lack of inflectional morphology (e.g., no conjugation for verbs and no declension for nouns in Chinese; Li, Jin, & Tan, 2004). Studies of crosslinguistic comparisons of early lexical development have also indicated patterns of distinct noun–verb acquisition as a function of the specific types of language. For example, many studies report that there is a preponderance of nouns compared to other categories of words in English-speaking children's early lexicon (Gentner, 1982; Bates et al., 1994), but there is no clear evidence of this noun bias in several Asian languages such as Chinese and Korean (Choi, 2000; Hao et al., 2015; Tardif, 1996, 2006; Xuan & Dollaghan, 2013).

Diversity of linguistic categories may be influenced by sociocultural factors (Malt & Majid, 2013) and emerge dynamically from the interaction between the learner and the learning environment (Elman, 2004; Li, 2009). The crosslinguistic differences in lexical and grammatical categories result in an absence of complete conceptual equivalence in the lexical vocabularies of different languages (Pavlenko, 2009). The degree to which a particular lexical concept varies across languages may result in different representations in the brain, which in turn may affect the success of cross-language brain decoding. Take grammatical categories for example. Previous studies have demonstrated that nouns and verbs are represented and processed in separate brain regions, with nouns engaging the temporal cortex more strongly and verbs the prefrontal areas more extensively (Shapiro & Caramazza, 2003; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011). However, in languages where the distinction between nouns and verbs is less clear grammatically (such as Chinese), they do not evoke distinct cortical activity as in English or other Indo-European languages (Li et al., 2004). Therefore, cross-linguistic differences in how grammatical class is expressed may affect cross-language brain decoding, particularly for decoding those words that are class-ambiguous in one language but not in the other language.

Some previous research has examined the influence of cross-language similarity on the neural representation by directly comparing between languages (Chan et al., 2008; Jeong et al., 2007; Kim et al., 2016; Tolentino & Tokowicz, 2011). Differences in brain activation of L1 and L2 were observed as a function of cross-language similarity in terms of word order (Jeong et al., 2007) and orthographic transparency (Kim et al., 2016). For example, Jeong et al. found that, in native Korean speakers who had learned two L2s (English and Japanese), there was greater neural similarity in the left inferior frontal gyrus, right superior temporal cortex and right cerebellum between Korean vs. Japanese sentence processing in comparison to Korean vs. English sentence processing. The results were interpreted as an effect of word-order similarity (e.g., Subject–Object–Verb in both Korean and Japanese as opposed to Subject–Verb–Object in English). In Kim et al.'s fMRI study (2016), Korean-Chinese-English trilinguals were tested using a word rhyming judgment task that tapped into the orthography-phonology mapping process. The distance of orthographic transparency was smaller between English and Korean than between Chinese and Korean. Results showed that Korean word processing activated largely overlapping brain areas as English word processing, whereas it led to substantial differential activations in bilateral frontal and temporal cortical areas as compared with Chinese word processing. A more recent study by Xu, Baldauf, Chang, Desimone, and Tan (2017) showed that despite overall overlapping brain activation, MVPA indicated distinct fine-grained patterns of neural representation between Chinese and English word processing. Therefore, it is important for future studies to take into consideration the degree of (dis)similarity between languages and examine what aspects of language properties may influence the success of cross-language brain decoding.

3.2. Age of acquisition (AoA) and proficiency

AoA and proficiency of the second language have been found to be among the most important variables underlying the neural representation of L1 and L2 in the bilingual brain (see Hernandez & Li, 2007; Perani & Abutalebi, 2005 for reviews). A recent meta-analysis by Cargnelutti, Tomasino, and Fabbro (2019) summarized 57 publications with regard to the brain activation patterns of L1 and L2 using activation likelihood estimation (ALE). They used age 6 as the AoA cutoff and found that late bilinguals (i.e., after age 6) consistently recruited more neural resources in the left inferior frontal gyrus and the posterior-medial frontal gyrus for processing L2 than processing L1, whereas this difference was not significant in early bilinguals (i.e., before age 6). The authors attributed the late bilinguals' additional neural activation in the prefrontal cortex to more effortful executive function required for processing in the L2. On the other hand, there is also evidence showing that earlier L2 AoA was associated with greater neural dissimilarity between L1 and L2. A recent study by Ou, Li, Yang, Wang, and Xu (2020) used representational similarity analysis (RSA) to quantify the degree of neural similarity between L1 and L2 processing. It was found that earlier AoA was associated with higher pattern dissimilarity between L1 and L2 in the left inferior gyrus and middle frontal gyrus. This AoA effect may be due to earlier bilinguals' greater neural plasticity in promoting language-specific neural computations for different languages, especially when learning two distant languages. This explanation was consistent with another recent study that used quantitative MRI (qMRI) in combination with fMRI techniques (Luo et al., 2019), in which early L2 AoA was associated with enhanced microstructural proliferation in Chinese-English bilinguals.

Many studies have shown that L2 proficiency is an equally important factor (see Hernandez & Li, 2007; Li, 2013 for reviews). In another meta-analysis using ALE, Sebastian, Laird, and Kiran (2011) focused on the neural activation modulated by L2 proficiency. They found that high-proficiency bilinguals showed more similarity in L1 and L2 brain representations compared to moderate/low-proficiency bilinguals, particularly in the left superior frontal gyrus and left middle temporal gyrus. Yang and Li (2019) also showed that the connectivity patterns in late L2 learners' brain network are moderated by L2 proficiency, among other abilities such as auditory pitch processing. There has also been recent work to identify the independent contributions of AoA vs. proficiency to neural representations and brain structures (see Nichols & Joanisse, 2016 for example), as well as convergent multimodal imaging evidence from resting-state, functional, and structural MRI investigations (Wang et al., 2020).

In Sheikh, Carreiras, and Soto (2019a) study, the influences of both AoA and L2 proficiency on cross-language decoding were examined. Although all participants were early bilinguals, their AoA of Spanish was earlier than that of Basque and they were also more proficient in Spanish than in Basque. It was hypothesized that more balanced bilinguals would exhibit increased cross-language decoding accuracy. They performed correlational analyses between cross-language decoding accuracy and the proficiency levels in Basque vs. Spanish (using proficiency difference scores). Results showed a tendency that the smaller the difference in proficiency between the bilingual's two languages, the greater the brain decoding accuracy in the left lateral temporal lobe, inferior frontal gyrus and dorsomedial prefrontal cortex. However, the authors noted that the results should be taken with caution because the study was not designed to explore inter-individual differences and that the results did not survive multiple comparison corrections.

3.3. Depth of language processing

Brain decoding within and across languages depends on the quality of the neuroimaging data obtained, which is in turn dependent on the participant's level of processing of the language stimuli. It is well known from classic memory theories that deeper, more elaborative, and richer semantic processing would lead to better memory (e.g., more successful retrieval) than shallow or surface-level processing of the same material (e.g., Craik & Lockhart, 1972). Some researchers propose that depth of processing during the processing task plays a critical role in the generalizability of semantic representations across languages.

Sheikh et al. (2019a) tested the role of depth of processing in early and proficient Spanish-Basque bilinguals, who were asked to perform shallow or deep semantic processing tasks. In the shallow processing task, the participants were asked to read and attend to the word, whereas in the deep processing task, the participants were asked to think about the characteristics of the living/non-living object it represented (e.g., shape, color). MVPA analyses showed that cross-language decoding of concepts was significant only in the context of deeper levels of processing, whereas the decoding of the word category within language was significant regardless of the level of depth of processing. This pattern of results indicates that deeper semantic processing (and the resulting brain patterns) is a necessary condition for cross-language decoding to be successful. The same researchers (Sheikh, Carreiras, & Soto, 2019b) further conducted a study to examine the influence of awareness of words on language decoding. Words were visually presented to Spanish-Basque bilinguals briefly with masks such that participants could not be consciously aware of the word. Participants were instructed to determine whether the word was animate or inanimate and rate how conscious they were about the word (e.g., 'didn't see anything' or 'saw the word clearly'). It was shown that fully or partially conscious conditions elicited above-chance decoding accuracy in more regions of interest (ROIs) than non-conscious conditions: when participants were unaware of the specific words due to rapid presentation and masking, above-chance decoding was found in limited regions for both Spanish and Basque; when participants were partially aware or fully conscious about the words, all ROIs showed above-chance within-language decoding accuracy. Further, their findings suggested that brain decoding across languages requires not only a deeper language processing but also conscious perception of language stimuli.

Given that depth of processing can significantly impact both cross-language and within-language decoding, many previous studies, in order to engage participants in deep processing, have presented the same word stimuli for multiple times during brain imaging (e.g., Buchweitz et al., 2012; Correia et al., 2014, 2015; Mitchell et al., 2008; Van de Putte et al., 2017, 2018; Zinszer et al., 2015). For example, in Mitchell et al.'s classic study, the participants viewed the same word/picture six times and their task was to think about the properties of the objects/concepts to which the stimuli refer when the stimuli were presented. The results could have been different if the stimuli were presented only one time, leading to shallow processing. Moreover, L1 and L2 words can differ in the processes and contexts within which learning takes place, which may influence the depth of processing in different languages. During L1 learning, children build direct relations between the words and the objects/concepts by integrating the perceptual and sensorimotor experiences from the environment, interacting with the objects and people and performing actions in it, whereas during L2 learning, the learners typically associate the words to an existing label in their native language (see Li & Jeong, 2020 for a review). The lack of embodied and social interaction during L2 learning may lead to a shallower and less elaborative processing of L2 words relative to L1, which may in turn affect the success of cross-language brain decoding in L2 vs. L1 stimuli (Jeong, Li, Suzuki, Kawashima, & Sugiura, 2020).

4. Beyond words: Sentence- and discourse-level decoding

So far, the cross-language decoding approach has been applied mostly to the word/concept level, as reviewed above. A few studies, with limited success, have extended the approach to the sentence and discourse levels. Yang, Wang, Bailer, Cherkassky, and Just (2017a) hypothesized that given the common neural substrates between languages, it is possible that the English-based model could also be applied to decode sentences in other languages. Therefore, the authors used the parameters from the English-based model of Wang, Cherkassky, and Just (2017), including brain locations associated with semantic dimensions, semantic features/thematic roles and trained model weights, to decode the Portuguese sentences that were translated from English (in Wang et al.'s study, each word in a sentence was encoded based on 42 semantic features, e.g., color/size/animacy, and 6 thematic roles, e.g., agent/patient/predicate). Both Portuguese monolinguals and Portuguese-English bilinguals were recruited to read the translated sentences in Portuguese. The decoding accuracy of the English-based model on Portuguese sentences was significantly above chance. Yang et al.'s findings also showed that the cross-language decoding accuracy did not vary as a function of whether the participants were monolinguals or bilinguals (0.67 for Portuguese monolinguals versus 0.66 for Portuguese-English bilinguals), suggesting that knowing English did not facilitate the decoding accuracy of Portuguese even though the decoding model was based on English language stimuli.

Yang, Wang, Bailer, Cherkassky, and Just (2017b) further extended their cross-language decoding across two languages to the decoding of sentences across three languages, i.e., English, Portuguese, and Chinese. Stimuli included sentences written in English, Portuguese and Chinese. English monolinguals, Portuguese monolinguals, Portuguese-English bilinguals and Mandarin-English bilinguals were asked to read sentences referring to both concrete and abstract concepts in their native language and think about the meanings of the sentences while in the scanner. The decoder was trained to map between sentences and the corresponding activations either in one language or in two languages (e.g., English and Chinese, English and Portuguese). The decoder trained on two languages (0.668 averaged across three language pairs) was more accurate when applied to the third language than the decoder trained on one language and then applied to either of the remaining languages (0.624 averaged across six one-to-one language pairs), and this advantage was especially pronounced with abstract concepts including social interactions and mental activity.

Yang et al. (2017b) used RSA to show that the similarity of neural representation of sentences between English and Portuguese was slightly higher than that between Portuguese and Chinese. This may suggest that the neural semantic space was more similar between similar languages (e.g., Portuguese and English) than between distant language (e.g., Portuguese and Chinese; see 3.1 on the role of cross-language similarity). However, the pairwise language decoding accuracy among the three languages did not differ significantly (e.g., decoding accuracy from English to Portuguese and from Chinese to Portuguese was 0.63 and 0.60, respectively). It is important to note that that the sample size of Yang et al.'s study was small (based on 7 participants from each language group) and the language experiences of the participants were not matched (e.g., all Chinese participants and 3 Portuguese participants were bilingual and the remainder were monolingual). These discrepancies may affect the decoding results and limit the generalization of their findings with regard to how language similarity may affect brain representations and decoding successes across languages.

Given the encouraging findings from brain decoding of words and sentences, researchers recently have also been interested in how discourse-level narrative stories are represented in the brain and whether cross-language brain decoding of narratives is possible. Dehghani et al. (2017) recruited English monolinguals, Chinese-English bilinguals, and Farsi-English bilinguals to perform in-scanner reading of translations of the same 40 stories in their native language (English, Chinese, Farsi, respectively). The 40 stories were chosen from an English corpus of over 20 million weblog story posts and were translated into Mandarin Chinese and Farsi. Doc2vec (Le & Mikolov, 2014) was used to model narrative-level semantic representations of the stimulus stories using a large weblog corpus for each language separately, and each of the stories in each language was represented as a 100-dimension semantic vector. Their result demonstrated that which specific story a participant was reading could be decoded using the fMRI signals from reading of these stories. Moreover, this decoding was successful even when the decoder based on a different language was applied to predict the story of the language the participant was reading. For example, the decoding accuracy of Chinese story was 0.55 when using the decoder of Farsi stories. The cross-language decoding accuracy was 0.564 averaged across all six language pairs. Searchlight-based MVPA indicated that informative clusters for cross-language and within-language story decoding were similar, mainly located in the default mode network including the posterior medial cortices, the medial prefrontal cortex, and the lateral parietal cortex. Although Dehghani et al.'s study provided initial evidence that cross-language decoding of narrative stories is possible, the accuracy remained quite low (0.564 on average), as compared with the accuracy seen with word or sentence-level decoding (typically in the 0.60-0.70 range). More work is needed in the future to examine discourse-level brain decoding across as well as within languages.

5. General discussion and future directions

5.1. Brain decoding in bilinguals vs. monolinguals

Most of the previous studies of cross-language decoding have recruited bilingual speakers so as to decode L2 from L1 and vice versa. Many of these studies reported above-chance cross-language decoding accuracy, but the accuracy was still lower than that for within-language decoding (Correia et al., 2015; Sheikh et al., 2019b). This is especially the case when more complex language materials such as stories are involved (e.g., Dehghani et al. reported accuracy of 0.564 on average). Given this situation, it is necessary to compare brain decoding based on bilingual participants' processing of two languages (L1 and L2) with brain decoding based on separate groups of monolinguals' processing of the same languages (both L1). It is possible that proficient bilinguals know the words in both languages well, and brain decoding may reflect the effect of association of the word equivalents. In addition, the socio-cultural experiences produce greater cross-language consistency within the same participant (the bilingual) than that between participants (two different monolinguals), such that neural representations of different languages may overlap to a greater extent in bilinguals than in two monolingual groups.

It is also important to note that bilingual imaging studies have shown that different languages are represented by both shared and distinct neural patterns, and cross-language decoding may fail when distinct neural representations occur in the first place. Given the previous findings on distinct neural patterns of representation (e.g., Li et al., 2004; Xu et al., 2017; Yang, Tan, & Li, 2011) and given our earlier discussion of the impact of language similarity on cross-language decoding accuracy, further systematic investigations into different languages are needed to more fully understand the extent to which cross-language decoding is feasible across similar and distinct languages (e.g., English vs. Chinese).

5.2. Beyond concrete concepts and single words

Studies of within-language decoding have shown that levels of concreteness may be an important factor for brain decoding: the higher the concreteness, the higher the decoding accuracy (e.g., Anderson, Murphy, & Poesio, 2014; Fernandino et al., 2015). Cross-language decoding studies have so far focused on decoding of a narrow range of concrete concepts (Buchweitz et al., 2012; Correia et al., 2014, 2015; Sheikh et al., 2019a, 2019b; Van de Putte et al., 2018; Zinszer et al., 2015), partly because concrete concepts are easier to test both in neuroimaging studies and computational modeling. The issue of whether cross-language decoding is feasible for abstract concepts remains largely unexplored. Unlike concrete concepts (e.g., "apple"), abstract concepts (e.g., "law") do not have specific external referents, and it is unclear how cross-language decoding of abstract concepts might differ as a function of the variables discussed above (i.e., language similarity, AoA, proficiency, depth of processing). Further studies should address whether the similarity of neural representations of abstract concepts are sufficient for successful decoding across languages. Such studies would also advance our understanding of the brain organization of different word categories for different languages.

Another important direction of research is to extend cross-language brain decoding from single word/concepts to higher levels of sentences and discourses. Moving beyond the single-word level to sentence- and text- levels should be an important direction for the neuroscience of language research in general (see Hagoort, 2019). Earlier we reviewed two previous studies focused on the sentence level (Yang et al., 2017a, 2017b) and one on the discourse level (Dehghani et al., 2017), but many more studies on these levels are needed given their importance. Brain decoding at the sentence and discourse levels can be more complex and challenging than single word level, since higher-level language stimuli involve additional complexity and variations of the words' thematic role, syntactic features, contextual information, and structured conceptual representation.

We hypothesize that discourse-level brain decoding may be more language-independent, given that, unlike meanings of single words, meanings of text-based paragraphs and narratives may not differ significantly across languages. For example, language similarity may be more important to affect the accuracy of brain decoding at the word level, given the language-specific properties of word stimuli. In Chinese, for example, words may show different semantic relations due to their orthographic structures and similarities (氵'water' directly provides the semantic clues to all words that share this radical such as 河 'river'). At the discourse level, however, such language-specific properties are absent. Discourse processing involves multiple levels of processing that are not language-specific, including coarse semantic processing, topical coherence monitoring, text integration, interpreting a protagonist's perspective, mental model building and so on (see Mason & Just, 2006 for a review). These processes activate a distributed network beyond the classical language areas (see Ferstl, Neumann, Bogler, & Von Cramon, 2008; Li & Clariana, 2019 for reviews). Extant studies indicate that neural activity associated with isolated words is primary driven by the properties of stimuli and accumulates information over relatively short time scales, while neural activity associated with narrative-level processing in high-order areas can accumulate information over longer periods of time (e.g., in areas such as precuneus, inferior frontal gyrus, medial frontal gyrus, temporoparietal cortex; Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Lerner, Honey, Silbert, & Hasson, 2011). Toneva and Wehbe (2019) recently showed that brain activities during reading of naturalistic texts in the frontal and parietal regions were mostly predicted by long-range contextual representations, which are distinct from brain activities predicted by word-level representations (Toneva & Wehbe, 2019).

5.3. From pattern classification to computational modeling

Previous studies of cross-language brain decoding have mainly been based on pattern classification methods such as the MVPA. Such methods have enabled us to predict patterns of brain activity for stimuli of a language using a training set of stimuli of another language and its associated neuroimaging data. An alternative approach for brain decoding is based on computational modeling (e.g., Mitchell et al., 2008), which enables us to test competing computational models and elucidate the extent to which the models are consistent with the stimulus representations in the brain. For example, Seyfried and Li (2020) used RSA to test context-dependent models (e.g., BERT, Devlin et al., 2018) and context-independent models (e.g., fastText, Bojanowski, Grave, Joulin, & Mikolov, 2017), and examined the degree to which these computational models represented information processing in the brain. Future studies should perform brain decoding with many competing models, each explaining a portion of the response-pattern variance (Kriegeskorte, 2011).

A related issue is how to capture and integrate language-specific and culture-specific properties into brain decoding models. The neural representations of specific concepts or relations between concepts can be affected by cultural factors associated with different languages. For example, discourse-level processing may involve background knowledge that is not part of the text content proper, but historical or cultural knowledge or information independent of the semantic content. Speakers of different languages and cultures have their own unique experience of the same word or concept due to different environments, including the concepts' cognitive and affective properties (Kuang, Li, Chen, Jin, & Chen, 2012). For example, the word "red" refers to a visual color in both English and Chinese, but in the Chinese culture it is also associated with celebration, enthusiasm, and happiness, while in Western cultures, it can be associated with debt, loss, and anguish. For these kinds of concepts, cross-language decoding accuracy may be influenced by the extent to which the bilingual materials share similar cognitive, affective, and emotionality properties (see Pavlenko, 2012 for discussion). At the same time, successes or failures of cross-language decoding could inform us of the significance of these nonlinguistic properties in the neural representation of language and concept. Previous work (Yang et al., 2016) has shown that certain regions in the brain, such as the right angular gyrus, may become particularly activated when processing information related to cultural background knowledge in the case of Chinese idioms. Such language-specific and culture-specific cases and their neural correlates suggest that cross-language decoding (from English to Chinese or vice versa) may be more challenging than we think.

6. Conclusions

Recent advances in machine learning research have opened up new ways for investigating the neural representation of language in the brain. Brain decoding has been an exciting and rapidly developing topic in this regard. Cross-language brain decoding has the potential to provide new insights into how our brain represents multiple languages. Our review of recent studies in cross-language brain decoding indicates that it is possible to decode semantic information across different languages from neuroimaging data, but there are also significant challenges to its success. Factors such as cross-language similarity, AoA/proficiency levels, depth of language processing may all affect the effectiveness of cross-language decoding. We expect to see continued progress in cross-language decoding, from a traditional focus on words and concrete concepts toward the use of naturalistic experimental tasks involving higher-level language processing (e.g., discourse processing). The cross-language decoding approach can also be applied to understand how cross-modal, cross-cultural, and other nonlinguistic cognitive and affective factors may influence neural representations of different languages. We need to design such studies with theoretical frameworks and hypotheses, which will in turn inform and contribute to the understanding of the cortical representations of different languages. Finally, future developments in both neuroimaging techniques and machine learning algorithms will allow us to capture highly detailed spatial and temporal information as language processing unfolds in real time, thereby enabling language and cross-language brain decoding with high fidelity.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Preparation of this article was made possible by grants from the National Science Foundation (BCS-1533625), a Faculty Startup Fund from the Hong Kong Polytechnic University, the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team (grant no. 2016ZT06S220), National Natural Science Foundation of China (31700951) and Shenzhen Basic Research Scheme (JCYJ20170412164259361).

References

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53*(1), 60–80.

Anderson, A. J., Murphy, B., & Poesio, M. (2014). Discriminating taxonomic categories and domains in mental simulations of concepts of varying concreteness. *Journal of Cognitive Neuroscience, 26*(3), 658–681.

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature, 568*(7753), 493–498.

Athanasopoulos, P. (2009). Cognitive representation of color in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition, 12*(1), 83–95.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J., et al. (1994). 1994: Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*, 85–123.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for. Computational Linguistics, 5*, 135–146.

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.

Cargnelutti, E., Tomasino, B., & Fabbro, F. (2019). Language brain representation in bilinguals with different age of appropriation of the second language: A meta-analysis of functional imaging studies. *Frontiers in Human Neuroscience, 13*, 154.

Chan, A., Luke, K., Li, P., Yip, V., Li, G., Weekes, B., & Tan, L. (2008). Neural correlates of nouns and verbs in early bilinguals. *Annals of the New York Academy of Sciences, 1145*, 30–40.

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language, 27*(1), 69–96.

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. *Journal of Neuroscience, 34*(1), 332–338.

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, 6*, 71.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology, 8*, 522.

Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., … Damasio, A. (2017). Decoding the neural representation of story meanings across languages. *Human Brain Mapping, 38*(12), 6096–6106.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings from NAACL-HLT 2018: The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana*.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences, 8*(7), 301–306.

Emmorey, K., Giezen, M. R., & Gollan, T. H. (2016). Psycholinguistic, cognitive, and neural implications of bimodal bilingualism. *Bilingualism: Language and Cognition, 19*(2), 223–242.

Fang, S. Y., Zinszer, B. D., Malt, B. C., & Li, P. (2016). Bilingual object naming: A connectionist model. *Frontiers in Psychology, 7*, 644.

Fernandino, L., Humphries, C. J., Seidenberg, M. S., Gross, W. L., Conant, L. L., & Binder, J. R. (2015). Predicting brain activation patterns associated with individual lexical concepts based on five sensory-motor attributes. *Neuropsychologia, 76*, 17–26.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29*(5), 581–593.

Gallant, J. (2016). Decoding language in the brain. *Research Features Magazine*, 11–13.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science, 366*(6461), 55–58.

Hao, M., Liu, Y., Shu, H., Xing, A., Jiang, Y., & Li, P. (2015). Developmental changes in early child lexicon in Mandarin Chinese. *Journal of Child Language, 42*, 505–537.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience, 28*(10), 2539–2550.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage, 62*(2), 852–855.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638.

Jeong, H., Li, P., Suzuki, W., Kawashima, R., & Sugiura, M. (2020). Neural mechanisms of language learning from social contexts. *Brain and Language* (in press).

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., … Kawashima, R. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping, 28*(3), 194–204.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). Boston naming test. Pro-ed.

Kim, S. Y., Qi, T., Feng, X., Ding, G., Liu, L., & Cao, F. (2016). How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean–Chinese–English trilinguals. *NeuroImage, 129*, 25–39.

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *Neuroimage, 56*(2), 411–421.

Kuang, H., Li, B., Chen, C., Jin, P., & Chen, X. (2012). Computation of bilingual word similarities by metaphorical properties. *Paper presented at the international conference on machine learning and cybernetics*.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196).

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience, 31*(8), 2906–2915.

Li, P. (2009). Lexical organization and competition in first and second languages: Computational and neural mechanisms. *Cognitive Science, 33*(4), 629–664.

Li, P. (2013). Neurolinguistic and Neurocomputational Models. In F. Grosjean, & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 214–238). New York, NY: John Wiley & Sons, Inc.

Li, P., & Clariana, R. B. (2019). Reading comprehension in L1 and L2: An integrative approach. *Journal of Neurolinguistics*, 94–105.

Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *npj Science of Learning, 5*(1), 1–9.

Li, P., Jin, Z., & Tan, L. H. (2004). Neural representations of nouns and verbs in Chinese: An fMRI study. *NeuroImage, 21*(4), 1533–1541.

Li, P., Tan, L.-H., Bates, E., & Tzeng, O. (2006). *The Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge University Press.

Luo, D., Kwok, V. P. Y., Liu, Q., Li, W., Yang, Y., Zhou, K., et al. (2019). Microstructural plasticity in the bilingual brain. *Brain and Language, 196*, 104654.

Macsweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: The neurobiology of sign language. *Trends in Cognitive Sciences, 12*(11), 432–440.

Malt, B. C., Jobe, R. L., Li, P., Pavlenko, A., & Ameel, E. (2016). What constrains simultaneous mastery of first and second language word use? *International Journal of Bilingualism, 20*(6), 684–699.

Malt, B., Li, P., Pavlenko, A., Zhu, H., & Ameel, E. (2015). Bidirectional lexical interaction in late immersed Mandarin-English bilinguals. *Journal of Memory and Language, 82*, 86–104.

Malt, B. C., & Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*, 583–597.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts. *Journal of Memory and Language, 40*, 230–262.

Malt, B. C., Sloman, S. A., & Gennari, S. (2003). Universality and language specificity in object naming. *Journal of Memory and Language, 49*(1), 20–42.

Mason, R. A., & Just, M. A. (2006). Neuroimaging contributions to the understanding of discourse processes. In M. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of Neuropsychology* (pp. 765–799). Amsterdam: Elsevier.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science, 320*(5880), 1191–1195.

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on second-language learning. *Neuroimage, 143*, 15–25.

Ou, J., Li, W., Yang, Y., Wang, N., & Xu, M. (2020). Earlier second language acquisition is associated with greater neural pattern dissimilarity between the first and second languages. *Brain and Language, 203*, 104740.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 125–160). United Kingdom: Multilingual Matters.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*(6), 405–428.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology, 15*(2), 202–206.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron, 72*(5), 692–697.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55.

Seyfried, F., & Li, P. (2020). Comparing sentence-based vs. word-based semantic space representations to brain responses. *Paper presented at the 30th Annual Conference of the Society for Text & Discourse (ST&D 2020 Online Meeting)*.

Sebastian, R., Laird, A. R., & Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Applied Psycholinguistics, 32*(4), 799–819.

Shapiro, K. A., & Caramazza, A. (2003). The representation of grammatical categories in the brain. *Trends in Cognitive Sciences, 7*(5), 201–206.

Sheikh, U. A., Carreiras, M. & Soto, D. (2019a). Neurocognitive mechanisms supporting the generalization of concepts across languages. PsyArXiv. doi:10.31234/osf.io/7g6dk.

Sheikh, U. A., Carreiras, M., & Soto, D. (2019b). Decoding the meaning of unconsciously processed words using fMRI-based MVPA. *Neuroimage, 191*, 430–440.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32*(3), 492.

Tardif, T. (2006). The importance of verbs in Chinese. In P. Li, L. H. Tan, E. Bates, & O. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics* (pp. 124–135). Cambridge, England: Cambridge University Press.

Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space, and time: A review of the role of cross-language similarity in L2 (morpho) syntactic processing as revealed by fMRI and ERP methods. *Studies in Second Language Acquisition, 33*(1), 91–125.

Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). In *Advances in Neural Information Processing Systems* (pp. 14928–14938).

Van de Putte, E., De Baene, W., Brass, M., & Duyck, W. (2017). Neural overlap of L1 and L2 semantic representations in speech: A decoding approach. *Neuroimage, 162*, 106–116.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews, 35*(3), 407–426.

Van de Putte, E., De Baene, W., Price, C. J., & Duyck, W. (2018). Neural overlap of L1 and L2 semantic representations across visual and auditory modalities: A decoding approach. *Neuropsychologia, 113*, 68–77.

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, 38*(10), 4865–4881.

Wang, R., Ke, S., Zhang, Q., Zhou, K., Li, P., & Yang, J. (2020). Functional and structural neuroplasticity associated with second language proficiency: An MRI study of Chinese-English bilinguals. *Journal of Neurolinguistics, 56*, 100940.

Yang, J., & Li, P. (2019). Mechanisms for auditory perception: A neurocognitive study of second language learning of Mandarin Chinese. *Brain Sciences, 9*(6), 139.

Yang, J., Li, P., Fang, X., Shu, H., Liu, Y., & Chen, L. (2016). Hemispheric involvement in the processing of Chinese idioms: An fMRI study. *Neuropsychologia, 87*, 12–24.

Yang, J., Tan, L. H., & Li, P. (2011). Lexical representation of nouns and verbs in the late bilingual brain. *Journal of Neurolinguistics, 24*(6), 674–682.

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017a). 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.

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017b). 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. *Brain and Language, 175*, 77–85.

Xu, M., Baldauf, D., Chang, C. Q., Desimone, R., & Tan, L. H. (2017). Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain. *Science Advances, 3*, e1603309.

Xuan, L., & Dollaghan, C. (2013). Language-specific noun bias: Evidence from bilingual children. *Journal of Child Language, 40*(5), 1057.

Zinszer, B., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R. D. (2015). You say potato, I say tǔdòu: How speakers of different languages share the same concept. *Paper presented at the 37th Annual Meeting of the Cognitive Science Society*.

Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R. D. S. (2016). Semantic structural alignment of neural representational spaces enables translation between English and Chinese words. *Journal of Cognitive Neuroscience, 28*(11), 1749–1759. https://doi.org/10.1162/jocn_a_01000.