Neural mechanisms of language learning from social contexts

Hyeonjeong Jeong a,b,c, Ping Li a, Wataru Suzuki d, Motoaki Sugiura b,e, Ryuta Kawashima b

a Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan
b Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan
c Department of Chinese and Bilingual Studies, Faculty of Humanities, the Hong Kong Polytechnic University, Hong Kong, China
d Miyagi University of Education, Sendai, Japan
e International Research Institute for Disaster Science, Tohoku University, Sendai, Japan

ARTICLE INFO

Keywords:
Social learning
Second language
Word encoding
Memory consolidation
fMRI

ABSTRACT

Humans learn languages in real-life situations by integrating multiple signals, including linguistic forms, their meanings, and the actions and intentions of speakers. However, little is known about the neural bases underlying the social learning of a second language (L2) in adults. In this study, 36 adults were asked to learn two sets of L2 spoken words through translation versus simulated social interactive videos (social learning). Brain activation during word learning was measured using fMRI. Greater activation was observed in the bilateral superior temporal sulcus, posterior middle temporal gyri, and right inferior parietal lobule during social learning as compared with translation learning. Furthermore, higher activity in the right temporal parietal junction, right hippocampus, and motor areas was observed during the initial stage of social learning, with the more successful performance being at the time of overnight testing. We argue that social learning may strengthen the link from new L2 forms to rich L2 semantic representations wherein memory properties are embodied, multimodal, and richly contextualized.

1. Introduction

One of the major research questions in second language (L2) studies is whether an enriched environment (e.g., real-life communicative contexts, interpersonal interaction) improves L2 skills, and if so, how this works. Indeed, first language (L1) learning or acquisition occurs in an interpersonal space where children integrate multiple signals, including linguistic forms (e.g., sounds), their meanings, and the actions and intentions of speakers (Bloom, 2000). This type of learning, operationally defined as social learning in this article, may enhance the understanding of not only the form and meaning of words but also about their function, that is, how to use these words in real-life social contexts.

In psychology, the method of encoding and processing during learning has long been believed to determine the quality and richness of semantic representation and to profoundly impact subsequent retrieval of learned materials (Craik & Lockhart, 1972; Morris, Bransford, & Franks, 1977; Tulving & Thomson, 1973). Many neuroimaging studies on memory also support the importance of encoding, showing that the degree of neural activity during encoding is a crucial factor in establishing memory representations (Davachi, 2006; Mandzia, Black, McAndrews, Grady, & Graham, 2004; Vaidya, Zhao, Desmond, & Gabrieli, 2002). Social learning as L1 acquisition may enrich semantic representations where memory properties are embodied, imageable, multimodal, and richly contextualized (Ellis, 2019), and supported by embodied cognition in the brain (Barsalou, 2008; Glenberg, Sato, & Cattaneo, 2008).

In contrast, most classroom L2 learners in countries where L2 is not used on a daily basis (e.g., most countries in Asia or Latin America) tend to use translation to acquire a new word, or associate a new word with its translated L1 information. This kind of learning, defined as translation learning in this article, may weakly connect word forms, meanings, and concepts, resulting in poor semantic representations. Although L2 researchers have documented the importance of social learning (Hall, 2019; Lantolf, 2006), little is known about the cognitive mechanisms involved in encoding (i.e., form-meaning mapping) through social learning in the brain. The ultimate goal of this research is to fill this gap (see Li & Jeong, 2020 for a recent review).

Despite the importance of social learning, several neuroimaging studies have focused only on traditional research paradigms such as translation learning or L2-picture paired associative learning (Breitenstein et al., 2005; Raboyeau, Marcotte, Adrover-Roig, & Ansaldo, 2010; Yang, Gates, Molenaar, & Li, 2015). It has been reported that the medial
temporal lobe plays an important role in the initial stages of word learning. In particular, the left hippocampus has been regarded as a crucial area for acquiring new words (Breitenstein et al., 2005; Davis, Di Betta, Macdonald, & Gaskell, 2009). The left inferior partial lobe and superior temporal gyrus along with premotor areas are associated with the acquisition of auditory information (Golestani & Zatorre, 2004; Wong, Perrachione, & Parrish, 2007; Yang et al., 2015). Furthermore, the left inferior frontal gyrus (IFG) and left middle temporal cortex are involved in processing and retrieving of stored semantic information (Sadre & Wagner, 2002; Hickok & Poeppel, 2007; Raboyeau et al., 2010). In summary, cortical brain areas in the left fronto-parietal network and hippocampal-memory-related brain regions support L2 vocabulary acquisition (see Rodriguez-Fornells et al., 2009 for a review).

A few recent neuroimaging studies have shown that different cognitive demands during social learning may influence the development of the brain networks underlying L2 knowledge (Jeong et al., 2010; Legault, Fang, Lan, & Li, 2019; Mayer, Yildiz, Macedonia, & von Kriegstein, 2015; Verga & Kotz, 2019). Mayer et al. (2015) investigated whether embodied actions by learning L2 words enhance rich semantic representations in the brain. In their study, the participants were required to learn L2 words with self-producing gestures, pictures, and verbal information. After 5 days of training, the participants performed a translation task that required them to use the targeted words. It was found that the words learned with gestures produced greater activation in the posterior superior temporal sulcus (STS) and motor areas than words learned with pictures and verbal information (Macedonia, Repetto, Ischebeck, & Mueller, 2019). Recently, Legault et al. (2019) demonstrated the effect of rich perceptual and sensorimotor experiences in virtual reality-based (VR) learning on the brain. Intensive training of VR vocabulary learning was found to enhance the cortical thickness of the right inferior parietal lobule (IPL) compared to L2-picture associative learning. The researchers also observed that L2 learners who showed increased cortical thickness in the right IPL had better scores on a delayed retention test.

Focusing on real-life social contexts, Jeong et al. (2010) reported that retrieving words learned from social contexts or translation were processed in different parts of the brain. In their experiment, outside the functional magnetic resonance imaging (fMRI), participants were asked to memorize target L2 words through simulated videos including joint activity in real-life situations and through translation. After memorizing all the target words, they performed a retrieval task inside the MRI scanner. It was found that partially different brain areas were involved in retrieval processing, depending on how words had been learned. While the left middle frontal gyrus was associated with retrieving words learned from translation, the right supramarginal gyrus (SMG) was related to retrieval of words learned from social contexts. Furthermore, neural activity in the right SMG during social learning was similar to that of processing of L1 words already acquired through social interaction with others. These findings indicate that social learning may strengthen mappings between L2 word form and native-like semantic representation during encoding.

The above studies led Li and Jeong (2020) to formulate the social L2 learning (SL2) model, wherein the SMG and inferior parietal regions, especially in the right hemisphere, form part of the brain network supporting the social learning of language. However, most previous studies supporting the idea that social learning may boost better memory have examined neural activity associated with retrieval, but not encoding, of words (cf. Macedonia et al., 2019). Hence, we know little about the brain systems engaged in the actual form-meaning mapping (i.e., encoding) and the extent to which qualitative and quantitative engagement of brain systems during encoding affects the acquisition of semantic representations of L2 words. Recent research on memory suggests that deep-elaborative encoding (involving active discovery, multimodal information, social/emotional processing, etc.) promotes rapid cortical learning at the initial stage of encoding, and this cortical activation contributes to later long-term memory formation (Brodt et al., 2016; Hebscher, Wing, Ryan, & Gilboa, 2019; Schott et al., 2013). These findings led us to assume that cognitive operations underlying form-meaning mapping from social contexts must be different from those of translation learning, and this difference determines the quality and richness of L2 semantic representations. Assuming that learners infer the meaning of a word from social situations or a simulated video where a speaker uses a target word in a communicative situation, they may use sophisticated strategies to infer the meaning using various types of multimodal information, such as the referential intentions of the speaker, tone of voice, and contextual information. We hypothesize that the ability to infer meanings from social contexts during encoding may influence the learner’s ability to retain L2 knowledge and subsequently affect the degree of learning success in social contexts. Building on the study by Jeong et al. (2010), the present study aimed to investigate the neural mechanisms underlying form-meaning mapping (i.e., encoding) in which learners acquire new L2 words from social contexts and their effects on flexible application of the learned words into new contexts (i.e., retrieval). The second aim of this study was to examine whether individual neural activation at the initial stage of encoding could predict subsequent successful learning. These neural mechanisms were compared to those where learners encoded new L2 words based on their L1 translations. In the social learning condition, participants were asked to grasp the meaning of words by watching several different movie clips in which target words were used in different real-life situations. In contrast, in the L1 translation condition, written L1 translations of targeted L2 words were directly presented to the participants. To control the amount of information between social learning and translation learning, we added an L1 control condition where the participants’ L1 words were used in social or written texts. In this condition, participants did not need to acquire new meanings. To examine how brain functions change with the progression of learning, we compared the initial stage of encoding with the final stage. Further, this study tested how participants would retain words encoded from the social context and translation 1 day after the encoding session (i.e., overnight sleep). In the retention test, we tested whether participants could apply their newly acquired knowledge in both the translation test and new social contexts. The flexible application of L2 vocabulary knowledge in new contexts is critical for successful L2 acquisition of native-like competence. This retention test score is used as an index of individual differences in the flexible application of L2 knowledge. After overnight sleep, newly learned words are likely to be stored in long-term memory and processed similarly to existing words because of offline consolidation (Davis et al., 2009).

Given the above background, we developed two hypotheses regarding the neurocognitive patterns underlying L2 learning. First, partially different brain areas would be recruited during encoding depending on the type of learning, translation vs. social. Second, neural activation at the initial stage of encoding would contribute to later long-term memory formation. Consistent with previous studies of word learning, language-related brain areas in the left hemisphere may be commonly involved in both types of learning, during which learners analyze linguistic codes such as semantic and phonological information. In addition, during social learning through simulated videos, which include two people interacting and using a target word in a series of real-life situations, learners would analyze social signals and activate motor actions to connect form-meaning mappings (sound to meaning in our target word). Therefore, brain regions implicated in action perception, multimodal integration, and social cognition may play essential roles during social learning condition (Garfield, Peterson, & Perry, 2001; Legault et al., 2019). Based on existing literature on the neural basis of social learning, we hypothesized that the right SMG, IPL, and posterior STS, as multimodal association and action perception areas, may be activated to support L2 social learning (Legault et al., 2019; Li & Jeong, 2020; Mayer et al., 2015). Furthermore, the “Theory of Mind” (ToM) network, including the temporal parietal junction (TPJ) and medial
prefrontal cortex (mPFC), would serve as a candidate network for encoding words from social contexts. These ToM areas are known to be associated with inference of intention, understanding of contexts, and pragmatic knowledge in social communication (Ferstl, 2010; Hagoort, 2019; Saxe & Powell, 2006). Such cognitive processes in social reasoning outside the classic language network may be crucial for understanding word meanings and building conceptual knowledge of new words in real-life language acquisition. Altogether, our investigation will provide new insights into our understanding of the neural bases of social learning and how these bases are mediated by environmental factors (e.g., type of learning) and learner factors (e.g., individual differences).

2. Methods

2.1. Participants

The participants were 36 healthy right-handed native Japanese speakers who had never studied Korean as an L2. They were undergraduate and graduate students at Tohoku University in Sendai, Japan. Their average age was 20.31 (range, 19–25 years; standard deviation [SD], 1.56). They also had taken other foreign language classes (such as German, Spanish, and French), other than Korean, for one semester during their university education. None of them had ever spent more than 1 month staying and/or studying overseas. They had normal vision and no history of neurological or psychiatric illness. Handedness was evaluated using the Edinburgh Handedness Inventory (Oldfield, 1971). Written informed consent was obtained from each participant. This study was approved by the Ethical Committee of Tohoku University.

2.2. Experimental procedures

The whole experiment was planned over a 2-day period (Fig. 1). On Day 1, the participants were asked to learn 24 novel Korean spoken words in the following two ways: watching video clips where these words were used in various real-life communicative contexts (S), and listening to the target words with their written L1 translations (T). Each L2 word was presented with eight different video clips in a social context (S_{L2}) or audio recording of eight different speakers in the translation (T_{L2}) condition. To control the amount of information and repeated exposure between conditions, the L1 social context (S_{L1}) and L1 text (T_{L1}) conditions were included. A complete learning session lasted around 6 h, including breaks, for each participant. Brain activation was measured using fMRI at two time points: Time 1, in which participants were initially exposed to the target words, and Time 2 in which they completely memorized all the words. After one night of sleep (Day 2), when memory consolidation was enhanced, participants’ knowledge of the memorized L2 words was tested outside the MRI.

2.3. Stimuli

The target Korean words consisted of verbs, adjectives, and greeting words frequently used in daily life and pooled from a previous study (Jeong et al., 2010). These words were divided into two groups to ensure that the numbers of verbs, adjectives, and greeting expressions were equal across S_{L2} and T_{L2}. The target words were used in learning either from S_{L2} or from T_{L2} and counter-balanced across participants. For the S_{L2} learning condition, we created 12 scenarios for each word. Then, in the video clips, one or two of the 10 recruited Korean native speakers acted out by using the target word in one situation. For example, to express the target word “Dowajo,” meaning help me in English, an actor tried to move a heavy bag and asked another person for help. To create as natural a situation as possible, we made movies in various environments such as schools, offices, houses, parks, stations, trains, restaurants, and parking lots. Each movie lasted for 3 s. Part of the movie clips (five out of 10) was also included in a previous study (Jeong et al., 2010).

A total of 288 movie clips were prepared for the Korean words. We pre-tested 10 native Japanese speakers, who had never studied Korean and were similar to the main study’s participants, to select the movies that best expressed the meaning of the target word. These volunteers were asked to randomly watch 12 movie clips, write down the meaning and point out ambiguous movies from which the meanings were difficult to guess. Following this procedure, we selected 10 social context movie

![Fig. 1. Experimental overview over a 2-day period. During fMRI scanning, two L2 learning conditions (social learning: S_{L2}, translation learning T_{L2}) and two L1 control conditions were presented. The presentation order was counter-balanced across participants. On Day 2, all words learned in Day 1 were tested in both new social situations (social test) and translation with new voices (translation test). L1: first language; L2: second language.](image-url)
clips for each word whose meaning all participants correctly wrote (i.e., the other two clips were not unanimously agreed to be clear). Eight movies were used for the learning section on Day 1, and two movies used in the test of Day 2 as new social contexts. In addition, we prepared an incorrect movie in which the target word was not correctly used in the social context for the test item.

For the translation learning condition, audio files for the spoken words were prepared. A total of 11 native speakers of Korean uttered each of the 24 words aloud, and their voices were digitally recorded. Each recorded Korean word was presented for 3 s as one trial, with the L1 translation text in white on a black background. Eight audio files per word were used for the learning stimuli, and three files used for correct and incorrect trials during the Day 2 test.

For the control condition, we prepared Japanese social videos and audio files with written texts using the participants’ L1, Japanese (S1, and T1, respectively). We selected 24 Japanese words comparable to the Korean target words according to word types (i.e., verbs, adjectives, and greeting words) and the number of their types. These L1 words are frequently used in daily life and have different meanings relative to Korean target words. We created 10 Japanese movie clips for each L1 word and asked 10 native Japanese speakers to point out ambiguous movies from which the meanings were difficult to guess. Finally, eight movie clips were selected because all participants unanimously agreed that these movie clips were clear enough for each L1 word. Audio files of these words were also prepared with eight native Japanese speakers for the L1 text condition. Each recorded L1 word was presented with the written word in white on a black screen. L1 words were divided in two sets and counter-balanced between the S1, and T1, conditions across participants.

2.4. fMRI learning task

Before attending an fMRI learning session, the participants were asked to listen to the sounds of 24 Korean words three times. The participants were then told that the semantic information of these words would be presented either by a movie clip or with an L1 translation in the following fMRI learning session. During fMRI scanning at Time 1, the participants wore MRI-compatible noise-cancelling headphones (Opto-acoustics Ltd., Israel), which reduced subjective MRI scanning noise while correctly presenting auditory stimuli, and they viewed sequences of movie clips or written texts on the screen. Videos and texts were presented in a block design paradigm for five conditions (S2, T2, S1, T1, and non-word condition as filler). We blocked stimuli using the same word type and learning condition.

For the S2 condition, four movie clips per target word were presented to the participants. They were asked to guess the meaning of the target word in the video to learn the word. For the T2 condition, 12 spoken words recorded by four different speakers per target word with written L1 translations were presented. As the control condition, they were asked to watch L1 movies in S1, and listen to L1 spoken words with written L1 words in T1, and to just understand the video and audio stimuli in the L1 condition. In the filler condition, the participants listened to 24 Korean non-word sounds with a fixation cross on the screen. Altogether, the total number of blocks was 12 words in each condition (Fig. 1). One block lasted 12 s for each word and consisted of four movie clips or four auditory sounds with a written text. After watching each set of stimuli per word, participants were requested to push a button. This was done to ensure that they were not sleeping during the task. A 12-s resting baseline was interspersed between blocks. The participants were instructed to look at the central fixation cross during the baseline period. The order of presentation of conditions was randomized across participants. The duration of the fMRI experiment was 1465 s.

After fMRI scanning at Time 1, outside the MRI machine, the participants continued to learn the target words using the same stimuli and experimental program used at Time 1. In this learning session, the participants were given headphones, and asked to remember each target word. Each participant repeatedly took the entire experimental sequence, including four conditions. They were asked to repeat the entire program 12 times. The total learning time took approximately 4 h, including short breaks in between. The learning time was determined based on a previous study (Jeong et al., 2010).

Before fMRI scanning at Time 2, we asked the participants to attend a self-vocabulary test. In this test, the participants were required to listen to auditory words and check whether they knew the meaning. All participants reported that they had memorized the target words. Then, the participants attended a learning task inside the fMRI. At Time 2, the procedure and design were the same as those of Time 1, but novel video clips and audio sounds with different voices were used. All experiments were performed using presentation software (Neurobehavioral Systems, Berkeley, CA, USA).

2.5. Vocabulary test (Day 2)

One day after the learning session (Day 2), we administered the vocabulary test to measure how much vocabulary knowledge participants had retained. The participants were given headphones, and they viewed new movie clips on a computer during the test. The new movie clips were not presented in the learning session. In this test, the participants were asked to judge each movie clip in terms of whether the words were correctly used in given social contexts (social test), and whether the spoken words were correctly translated into their L1 (translation test). In the social test, the participants watched a short movie for 3 s via headphones and a computer screen, and pushed a button as soon as they knew the answer. In the translation test, which lasted maximally two seconds, each Korean spoken word was presented via headphones with written L1 translations on the screen. The participants were asked to push a button as quickly as possible.

All Korean words learned in the social context (S1, L1) and L1 translation (T1, L1) were tested both in social and translation tests. Each word consisted of two correct trials and one incorrect trial in both social and translation test conditions. All items were randomly presented to participants in each test. Thus, a total of 144 test items were used in the social test and translation tests (24 words × 3 items in social and translation tests). Test order was counter-balanced across participants. Reaction time and responses were digitally recorded using a computer. Correlations between test scores and brain activity during encoding were computed to examine individual differences.

2.6. Data acquisition

A time-course series of 977 volumes each in Time 1 and Time 2 was acquired using T2*-weighted gradient echo-planar-imaging sequences and a 3-Tesla MR imager (Achieva Quasar Dual, Philips Medical Systems, Best, The Netherlands). Each volume consisted of 25 axial slices covering the entire cerebrum and cerebellum (echo time = 30 ms; flip angle = 70°; slice thickness = 5 mm; no slice gap; field of view = 192 mm; 64 × 64 matrix; voxel dimension = 3.0 × 3.0 × 3.0 mm). Repetition time was 1500 ms. In addition, high-resolution T1-weighted structural MR images (TR = shortest; TE = shortest; FOV = 240 mm; matrix size = 240 × 240; 162 sagittal slices of 1-mm thickness) were also acquired. The following preprocessing procedures were performed using Statistical Parametric Mapping (SPM12) software (Wellcome Centre for Human Neuroimaging, London, UK) and MATLAB (MathWorks, Natick, MA, USA): adjustment of acquisition timing across slices, correction for head motion, co-registration to the anatomical image, spatial normalization using the anatomical image and the MNI template, and smoothing using a Gaussian kernel with a full width at a half-maximum of 8 mm. We excluded two participants from the statistical analyses as they had shown >3 mm excessive motion within the scanner. Thus, imaging data from 34 participants were included in the final analysis.
2.7. fMRI analysis

Conventional first-level (within-participant) and second-level (between-participants) analyses were performed using SPM12. In first-level analysis, a voxel-by-voxel multiple regression analysis of expected signal changes for each condition, constructed using the hemodynamic response function provided by SPM12, was applied to the preprocessed images for each participant. The degree of activation was estimated using a multi-session design matrix modeling the five conditions (S\_T\_1, T\_1, S\_1, T\_1, and filler) in each session (Time 1 and Time 2), and the six movement parameters in each session were computed at the realignment stage. Contrast images for Time 1 and Time 2 in each S and T condition, and L2 and L1 conditions at each time point were created for each participant.

Statistical inference on contrasts of parameter estimates was then performed using a second-level between-participants (random effects) model with one-sample t-tests. The statistical threshold in the voxel-by-voxel analysis, assuming a search area of the whole brain, was set at p < 0.001 for the initial height threshold and then corrected to p < 0.05 for multiple comparisons using cluster size. We also used a voxel-wise FWE-threshold of p < 0.05 to report the peak voxels that survived this criterion. After completing the estimation of main contrast (a whole brain voxel-by-voxel analysis), an inclusive masking procedure was applied to retain the voxels that also reached the level of significance in the masking contrast (mask threshold of p < 0.001 for height). The search area for the correction of multiple comparisons was not affected by these inclusive masks. To illustrate the activation profile in the observed brain area, we extracted parameter estimates in the four conditions for each participant using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002).

2.7.1. Effect of learning progress: Comparison between Time 1 and Time 2

To examine brain activation induced by the learning progress of L2 words in each type of learning, Time 1 was compared separately with Time 2 for each learning condition. We assumed that activation of brain areas relevant to learning (i.e., form-meaning mapping) recruited most at Time 1 would drastically decrease at Time 2 as learning progresses. To confirm our assumption, the contrasts of \([\text{Time } 1\_S\_2 > \text{Time } 2\_S\_1]\) for S\_2 and \([\text{Time } 1\_T\_2 > \text{Time } 2\_T\_1]\) for T\_2 were tested. To limit brain areas which also showed higher activation under L2 than L1 conditions (i.e., brain areas relevant to only L2 learning) at Time 1, the contrast of \([\text{Time } 1\_S\_2 > \text{Time } 1\_S\_1]\) or \([\text{Time } 1\_T\_2 > \text{Time } 1\_T\_1]\) was applied as an inclusive mask for S\_2 and T\_2, respectively. Additionally, to test whether there were any higher activation areas at Time 2 than Time 1 in each learning type, the contrasts of \([\text{Time } 2\_S\_2 > \text{Time } 1\_S\_2]\) for S\_2 and \([\text{Time } 2\_T\_2 > \text{Time } 1\_T\_2]\) were tested with the inclusive mask of \([\text{Time } 2\_S\_2 > \text{Time } 2\_S\_1]\) or \([\text{Time } 2\_T\_2 > \text{Time } 2\_T\_1]\), respectively. Inclusive masking was also applied to limit brain areas showing higher activation in L2 than in L1 at Time 2.

2.7.2. Effect of learning type: Comparison between S\_2 and T\_2

To identify the brain areas involved in different learning types (S\_2 and T\_2) during the initial stage of learning (i.e., form-meaning mapping), the contrasts of \([S\_2 > T\_2]\) or \([S\_1 > T\_1]\) and \([T\_2 > S\_2]\) or \([T\_1 > S\_1]\) at Time 1 were tested for the effect of social learning (S\_2) or L1 translation (T\_2), respectively. To limit brain areas showing activation in learning progress (i.e., higher activation at Time 1 than Time 2), the contrast of \([\text{Time } 1\_S\_2 > \text{Time } 2\_S\_2]\) or \([\text{Time } 1\_T\_2 > \text{Time } 2\_T\_2]\) was applied as an inclusive mask for the effect of social learning (S\_2) or translation learning (T\_2), respectively. The activation depicted by these analyses reflected differential brain activation between types of learning during the initial stage of L2 encoding. However, this activation did not reflect different amounts of visual information between S and T conditions.

2.7.3. Effect of individual differences

To examine the effect of subsequent successful retrieval on brain activity during the initial stage of encoding, we performed a voxel-wise single regression analysis with retention scores as an independent variable and brain activation as a dependent variable. The contrast of \([\text{Time } 1\_S\_2 > \text{Time } 2\_S\_2]\) and \([\text{Time } 1\_T\_2 > \text{Time } 2\_T\_2]\) was used for activation during the learning of S\_2 and T\_2, respectively. This retention score was measured by the vocabulary test on Day 2 after an overnight sleep consolidation.

3. Results

3.1. Behavioral data

Table 1 shows the mean percentage of correct responses and reaction time (RT) obtained from the vocabulary test administered on Day 2. Note that this test assessed whether both S\_2 and T\_2 words were correctly used in novel situations (social test) and whether they were correctly translated into their L1 (translation test). A two-way repeated-measures ANOVA was conducted to evaluate the main effects of learning type and test type, and the learning by test interaction on two dependent variables: accuracy and RT.

The main effects of learning type and learning by test interaction were found to be significant for both accuracy \(F(1, 33) = 7.2, p < 0.05\), and \(F(1, 33) = 17.89, p < .001, respectively\) and RT \(F(1, 33) = 17.31, p < .001, and F(1, 33) = 39.85, p < .001, respectively\). Words encoded from the S\_2 condition showed significantly higher accuracy rates and faster RTs than words encoded from the T\_2 condition.

Paired-samples t-tests were conducted to follow up on significant interactions. The accuracy rate of words encoded in the T\_2 condition was significantly lower than that of words in the S\_2 condition in the social test \(t(33) = 3.64, p < .001\), but no difference was found in the translation test. The RT was also significantly longer for words encoded from the T\_2 condition than those encoded from the S\_2 condition in the social test \(t(33) = -5.40, p < .001\), but no difference was found in the translation test. In other words, the words encoded from social contexts (S\_2) led to more accurate application of learned words to novel contexts than did those encoded from translations. In contrast, targeted words encoded with L1 translation (T\_1) dramatically deteriorated in the social test.

3.2. Imaging data

3.2.1. Brain areas induced by learning progress in each learning type

We confirmed the assumption that brain areas relevant to learning recruited the most at Time 1 would decrease at Time 2 as learning progresses. The brain regions showing the learning effect between Time 1 and Time 2 in each learning condition are shown in Table 2 and Fig. 2. The S\_2 condition at Time 1 produced significantly greater activation than its counterpart at Time 2 in the following brain areas: the bilateral IFG, the bilateral posterior middle temporal gyrus (pMTG), right STS, and supplementary motor area (SMA). The T\_2 condition at Time 1 also induced significantly greater activation in the left IFG and SMA than its counterpart at Time 2. In both social learning and L1 translation

| Table 1 |
| --- |
| Mean accuracy and RT. |
| N = 34 | Test type | S\_2 words | T\_2 words |
| Accuracy | S | 90.44 (7.26) | 82.92 (11.39) |
| | T | 93.80 (4.36) | 94.03 (5.43) |
| Reaction time | S | 3.29 (0.65) | 3.70 (0.79) |
| | T | 1.21 (0.30) | 1.17 (0.38) |

Mean accuracy (percentage of correct responses) and RT (seconds) are shown. SDs shown in parentheses. S: social test; T: translation test; RT: reaction time; SD: standard deviations
The threshold was set at the cluster-level threshold of \( p < .05 \) after FWE correction across the whole brain. * \( p < .05 \) after FWE correction across the whole brain. 

**3.2.2. Brain areas associated with an initial stage of learning: Comparison between \( S_{L2} \) and \( T_{L2} \)**

Comparisons between \( S_{L2} \) and \( T_{L2} \) produced significantly greater activation in the bilateral STS, pMTG, and the right IPL for the \( S_{L2} \) than \( T_{L2} \) conditions (see Table 3 and Fig. 3). However, significantly greater activation was not observed for \( T_{L2} \) than for \( S_{L2} \).

**3.2.3. Brain areas associated with individual differences for successful retrieval**

A voxel-level whole brain analysis revealed significant positive correlations between vocabulary scores of \( S_{L2} \) words and brain activity during social learning (contrast \( [S_{L2} > T_{L2}] \)) in the following areas: the right TPJ, the right hippocampus, and the left central and precentral motor areas (Table 4 and Fig. 4). The greater activation in these areas was produced during the initial stage of social learning, the higher vocabulary scores on Day 2 (after one night of sleep consolidation) were obtained. In contrast, there was no such correlation in any brain area in the \( T_{L2} \) condition.

**Table 2**

| Brain areas involved in learning progress between \( Time 1 \) and \( Time 2 \). |
|-------------------------------------------------------------|
| Brain areas \( x, y, z \) T-value cluster size |
| **Time 1.** \( S_{L2} > T_{L2} \) \( S_{L2} \) | Inclusive mask \( Time 1 \) \([S_{L2} > T_{L2}]\) |
| Left superior temporal gyrus | –60, –46, 16* | 6.70 | 642 |
| Left posterior middle temporal gyrus | –42, –43, 1* | 6.39 |
| Left inferior frontal gyrus | –60, –43, 4* | 5.73 |
| Left inferior frontal gyrus | –52, 7* | 5.78 | 427 |
| Left supplementary motor area | –8, –17, 16* | 5.15 |
| Right superior temporal sulcus | 59, –40, 10* | 5.91 | 579 |
| Right posterior middle temporal gyrus | 68, –52, 7* | 5.99 |
| Right inferior frontal gyrus | 52, 23, 7* | 6.28 | 353 |
| Right inferior parietal lobule | 42, 17, 13* | 5.21 |

For each area, the coordinates \((x, y, z)\) of the activation peak in MNI space, peak \( T \)-value, and size of the activated cluster in number \((k)\) of voxels \((2 \times 2 \times 2 \text{ mm}^3)\) are shown for all subjects \((n = 34)\). The threshold was set at the cluster-level threshold of \( p < .05 \) after FWE correction across the whole brain. * \( p < .05 \) after FWE correction (voxel-level).

Table 3

| Brain areas associated with an initial stage of learning: Comparison between social learning and translation learning. |
|---------------------------------------------------------------|
| Brain areas \( x, y, z \) T-value cluster size |
| **Time 1.** \( [T_{L1} > T_{L2}] \) \( [T_{L1} > T_{L2}] \) | Inclusive mask \( Time 1 \) \([T_{L2} > T_{L1}]\) |
| Left superior temporal sulcus | –42, –31, 4* | 5.95 |
| Left posterior middle temporal gyrus | –48, –16, 2* | 5.03 | 364 |
| Right superior temporal sulcus | 51, –19, –2* | 5.51 | 502 |
| Right posterior middle temporal gyrus | 45, –31, 10* | 5.34 |
| Right posterior parietal lobule | 60, –24, 4* | 4.98 | 204 |
| Right inferior parietal lobule | 30, –34, 43 | 4.56 |
| Right inferior parietal lobule | 27, –49, 61 | 4.37 | 255 |

For each area, the coordinates \((x, y, z)\) of the activation peak in MNI space, peak \( T \)-value, and size of the activated cluster in number \((k)\) of voxels \((2 \times 2 \times 2 \text{ mm}^3)\) are shown for all subjects \((n = 34)\). The threshold was set at the cluster-level threshold of \( p < .05 \) after FWE correction across the whole brain. * \( p < .05 \) after FWE correction (voxel-level).

**Brain areas involved in learning progress between \( Time 1 \) and \( Time 2 \)**

| a. Social learning |
|-------------------|
| b. Translation learning |

Fig. 2. Brain areas showing higher activation at \( Time 1 \) than in \( Time 2 \) in each learning condition (in the right MTG, STS and IFG for social learning and in left IFG and SMA for both learning conditions). The threshold was set at the cluster-level threshold of \( p < .05 \) after FWE correction across the whole brain. IFG: inferior frontal gyrus, STS: superior temporal sulcus, MTG: middle temporal gyrus; SMA: supplementary motor area.
This study aimed to investigate the neurocognitive mechanisms involved in the learning of new L2 words from social contexts and the effect of individual differences underlying these mechanisms on subsequent retrieval of new words. To do so, we compared the participants’ brain activation patterns during social learning of L2 with those during translation learning, a traditional but still popular L2-learning method. The left IFG and SMA were found to be activated during learning in both social learning and L1 translation conditions. This finding is consistent with that of many previous neuroimaging studies that investigated the effects of traditional word learning methods. Our results further showed that social learning uniquely induced neural activation in the bilateral STS, pMTG, and part of the right IPL during learning. The higher activation in these areas led to enhanced performance in the delayed vocabulary test where the participants applied target words to new situations. That is, the words encoded from social contexts showed better subsequent retrieval performance than those encoded from L1 translation. Furthermore, successful learners engaged in social learning were more likely to recruit the right hippocampus, right TPJ (including both SMA and AG), and motor areas during the initial stage of learning than less successful learners. Taken together, these findings indicate that social learning of L2 words may result in stronger activation in brain regions implicated in social, affective, and perception-action-related processing, which can boost rich semantic representation and flexible application of new words in new contexts. Below, we discuss the implications of our study’s findings.

### 4. Discussion

This study aimed to investigate the neurocognitive mechanisms involved in the learning of new L2 words from social contexts and the effect of individual differences underlying these mechanisms on subsequent retrieval of new words. To do so, we compared the participants’ brain activation patterns during social learning of L2 with those during translation learning, a traditional but still popular L2-learning method. The left IFG and SMA were found to be activated during learning in both social learning and L1 translation conditions. This finding is consistent with that of many previous neuroimaging studies that investigated the effects of traditional word learning methods. Our results further showed that social learning uniquely induced neural activation in the bilateral STS, pMTG, and part of the right IPL during learning. The higher activation in these areas led to enhanced performance in the delayed vocabulary test where the participants applied target words to new situations. That is, the words encoded from social contexts showed better subsequent retrieval performance than those encoded from L1 translation. Furthermore, successful learners engaged in social learning were more likely to recruit the right hippocampus, right TPJ (including both SMA and AG), and motor areas during the initial stage of learning than less successful learners. Taken together, these findings indicate that social learning of L2 words may result in stronger activation in brain regions implicated in social, affective, and perception-action-related processing, which can boost rich semantic representation and flexible application of new words in new contexts. Below, we discuss the implications of our study’s findings.

### 4.1. Brain network involved in social learning

Our data are consistent with the new SL2 framework by Li and Jeong (2020), according to which a brain network involving key regions is recruited to support the social learning of second languages. According to the SL2 framework, social learning draws on the link between the linguistic form and embodied system associated with both action perception and social-affective processes during encoding (i.e., form-meaning mappings). Even during simulated social interactions (e.g., video-based learning), learners would recruit broad brain regions to process multiple perceptual, action-related, social, and emotional cues to add meaning. These processes have the advantage of strengthening the associations between the new L2 words and embodied semantic representations in the brain in comparison with simple associative learning.
The higher involvement of the bilateral STS, pMTG, and right IPL in social learning than translation learning may reflect that these areas are involved in the analyses of social interaction signals and motor actions during learning. These analyses are essential cognitive processes for linking novel word forms to conceptual knowledge in language acquisition. In the current study, we used simulated videos that included two people interacting and using a target word with each other in a series of situations. By observing how target words were used in those situations, the participants were able to infer word meanings. To grasp the meanings of words, participants were likely to analyze relevant information such as the speaker’s intention, emotion, action, and salient features in the environment, and integrate these signals during form-meaning mapping.

In the present study, we observed broad areas of activation in the bilateral STS during social learning and argued that this brain region is associated with social perception during the encoding of L2 words. It has been reported that the STS is involved in the processing of various types of social information, ranging from the physical perception of voice, face, and biological motion to higher mental processes (ToM) and language comprehension (Beauchamp, 2015; Deen, Koldewyn, Kanwisher, & Saxe, 2015; Saxe & Powell, 2006). Deen et al. (2015) investigated the functional organization of the STS by measuring its responses to a range of social and linguistic stimuli. Their results showed not only a regular anterior-posterior organization in the STS for different social stimuli but also middle and posterior overlapping regions responsive to face, biological motion, language, and ToM. In the present study, we found two key patterns: (1) activation in the middle and posterior parts of the STS for the social learning condition but not the translation condition, and (2) higher activation in these areas in the right hemisphere than in the left. Taken together, the STS may play an essential role in analyzing and integrating the social signals that provide critical cues to learners when acquiring the meaning of L2 words.

Greater involvement of the bilateral STS and right IPL in social learning in the current study are consistent with those of previous neuroimaging studies on L2 vocabulary learning with gestures (Mayer et al., 2015) and on L2 vocabulary learning with a virtual environment (VE) (Legault et al., 2019). Mayer et al. (2015) investigated the effect of enrichment of embodied action during vocabulary learning. Words learned with gestures led to greater activation in the posterior STS along with motor areas than words learned with pictures and verbal information. IPL, especially in the right hemisphere, has long been regarded as a hub for L2 vocabulary acquisition (Della Rosa et al., 2013; Mechelli et al., 2004). Legault et al. (2019) also observed that L2 vocabulary learning through VE changed brain structures (i.e., increased gray-matter thickness) in the right IPL. They interpreted this finding as evidence that VE enhances rich perceptual and sensorimotor experiences during learning. Furthermore, the STS and right IPL observed by Mayer et al. (2015) and Legault et al. (2019), respectively, were highly sensitive to the subsequent performance of L2 words. These brain areas act to promote language learning with enriched sensorimotor experiences, not simply reflecting the reactivation effect of enactment. In the present study, we found similar effects in these areas, even when employing simulated social videos. The learners in the current study may have simulated embodied actions and social interaction in the brain during encoding.

Social learning may strengthen richer conceptual and semantic representations of L2 words than translation learning. The pMTG is considered to play a role in both controlled retrieval of conceptual knowledge and comprehension of events, relations, and actions (Bedny, Dravida, & Saxe, 2013; Binney & Ramsey, 2020; Davey et al., 2016). In the current study, we used verbs, greeting words, and adjectives that express an event in social situations, and observed an activation in the pMTG during social learning. Thus, participants may have processed concepts of events or actions underlying a target word’s meaning by watching a series of social situations in the videos. In the L1 translation condition where the same target words were used, however, we did not detect any activation in the pMTG.

An alternative explanation for activation in the pMTG during social learning may strengthen richer conceptual and semantic representations of L2 words than translation learning. The pMTG is considered to play a role in both controlled retrieval of conceptual knowledge and comprehension of events, relations, and actions (Bedny, Dravida, & Saxe, 2013; Binney & Ramsey, 2020; Davey et al., 2016). In the current study, we used verbs, greeting words, and adjectives that express an event in social situations, and observed an activation in the pMTG during social learning. Thus, participants may have processed concepts of events or actions underlying a target word’s meaning by watching a series of social situations in the videos. In the L1 translation condition where the same target words were used, however, we did not detect any activation in the pMTG.

 Greater involvement of the bilateral STS and right IPL in social learning in the current study are consistent with those of previous
learning is that this area may be associated with the encoding of word usage in social interactive contexts. The pMTG has been suggested to play an important role in the successful retrieval of words in communicative contexts (Grande et al., 2012; Jeong et al., 2015). Jeong et al. (2015), using fMRI, examined neural correlates of L2 communicative speech production towards others. The L2 learners in their study who acquired better communicative skills recruited the left pMTG more than less skilled L2 learners when producing contextually appropriate communicative speech.

4.2. Memory consolidation in social learning of L2

The aforementioned explanations of the impact of social learning on brain activation are supported by our behavioral data. According to the delayed retention test, where target words were used in new social contexts, the L2 words encoded from social contexts (i.e., simulated video) were found to be more effective than those encoded from L1 translation (Table 1). In this test, participants were asked to judge whether words were correctly used in novel situations (social test) or whether they were correctly translated into their L1 (translation test). All words encoded from the social condition and translation condition were tested in both social and translation tests. The words encoded from social contexts showed better accuracy in the social test than those encoded by translations. Those encoded from social contexts had better performance scores in the translation test, even when encoding and retrieval operations were different. In contrast, targeted words encoded with L1 translation scored dramatically worse in the social test.

These findings are consistent with the level of processing theory (Craik & Lockhart, 1972) and encoding specificity theory (Tulving & Thomson, 1973). These theories suggest that more elaborate semantic processing during encoding would lead to more successful retrieval than surface-level processing of the same items. In the current study, the social learning condition may have promoted elaborative processing in terms of overall larger cognitive efforts elicited by the processing of multiple social, perceptual, and action cues, and inferring word meanings from social contexts. Inference may have been taken place not only in each situation (a single video), but also across a series of situations (multiple videos). Such cognitive processes may have enhanced deeper semantic representations of L2 words and allowed for flexible application of L2 vocabulary knowledge into new social contexts. Our claim is supported by the imaging results that larger cognitive efforts during learning conditions. We created video files for social learning and audio files with written texts for translation learning with the same duration for each word. The participants in these two conditions were required to watch and listen to the targets within an equal time. Accordingly, we argue that new words encoded from social contexts can facilitate memory consolidation (i.e., rich semantic memory) after overnight sleep than words encoded by translation. Consequently, enriched semantic memory through social learning can be successfully applied to new social contexts and retained with more accuracy.

4.3. Successful social learning of L2 and individual differences

In the social learning condition, there were positive correlations between the delayed retention test scores and brain activity in the right TPJ, left post and precentral areas, and right hippocampus. In other words, when learning targeted L2 words, the participants who recruited both areas showed more successful learning and better memory performance than those who did not. It is important to note that such individual differences were not observed in the L1 translation condition. These results support our explanation above that social learning of L2 words enhanced elaborate cognitive processing and thus recruited broader brain areas relevant to social learning, hence leading to more successful learning. Processing multiple social, emotional, perceptual, and action-related meaning cues that co-occur during encoding may have strengthened direct mappings between novel L2 word forms and embodied semantic representations. This would allow for faster access and better accuracy in the flexible application of L2 knowledge in new contexts. Successful learning can be implemented in the brain networks underlying multimodal integration, social reasoning, motor simulation, and long-term memory (Li & Jeong, 2020).

The right TPJ has long been recognized not only as multimodal association area that integrate multisensory information (Macaluso & Driver, 2003), but also as a ToM area that plays an important role in social reasoning, such as thinking about other people’s beliefs, emotions, and intentions (Prat, Mason, & Just, 2011; Saxe & Powell, 2006). We assume that both functions are important in social learning. To understand the meaning of target words by watching real-life videos comprising verbal and nonverbal cues and contextual information, the learner critically needs social reasoning based on the integrated multimodal information. Although there are limited neuroimaging studies of L2 social learning, most have consistently reported activation in the right TPJ or adjacent areas (Jeong et al., 2010; Legault et al., 2019; Verga & Kotz, 2019; Liu et al., 2020). Liu et al. (2020) investigated the retrieval effect of multimodal learning using natural real-life videos. Stronger activation in the right TPJ was observed when the participants retrieved bimodal naturalistic events, such as watching a video, than when they retrieved unimodal events, such as only reading text or listening to a story. Our result is also supported by previous studies on memory (Schott et al., 2013), which demonstrated strong involvement of the right TPJ along with other ToM areas (i.e., medial PPC) when participants are engaged in deep encoding tasks (i.e., rate the pleasantness of a word’s meaning) compared to shallow encoding tasks (i.e., judging the number of syllables of a word). Indeed, the ToM network including the right TPJ has been reported in many neuroimaging studies on pragmatic language processing (Prat et al., 2011), narrative story comprehension (Ferstl, Neumann, Bogler, & von Cramon, 2006), and communicative speech acts (Jeong et al., 2015; Sussa et al., 2007; van Ackeren, Casasanto, Bekkering, Hagoort, & Rueschemeyer, 2012). In these types of language processing, people are required to infer the meanings of texts and utterances. The precise role of social reasoning in the development of L2 proficiency should be investigated in future studies.
The left post and precentral areas activated in the present study may be related to motor simulation while the participants watched simulated social videos. Compatible with theories of embodied cognition (Barsalou, 2008; Glenberg et al., 2008) and encoding specificity (Tulving & Thomson, 1973), recruitment of sensorimotor areas with rich social input during encoding can boost rich semantic representation of L2 words and better memory. Along with the right TPJ and right hippocampus, the left sensory motor areas may support a process in which learners build an embodied semantic representation of L2 words. These patterns are also consistent with a recent study by Zhang, Yang, Wang, and Li (2020) that directly compared L1 and L2 embodied semantic representations. Zhang et al. showed that the neural networks recruited during embodied semantic processing differed significantly between L1 and L2 speakers. Their L1 speakers engaged in a more integrated brain network connecting key areas for language and sensorimotor integration (e.g., the SMA), whereas L2 speakers failed to activate the necessary sensorimotor information and recruited a less integrated embodied system. Consistent with these patterns, the current data suggest that when L2 is acquired through social learning, the learner can build an embodied representation that resembles L1 representation more closely by engaging the brain’s sensorimotor regions.

A sizable neuroimaging literature on memory has reported that enhanced activation in the hippocampus is associated with successful encoding of various items and words (Berens, Horst, & Bird, 2018; Breitenstein et al., 2005; Davis et al., 2009; Mestre-Misse, Munte, & Rodriguez-Fornells, 2009; Takahshima, Bakker, Van Helle, & Janzen, 2014). Berens et al. (2018) provided evidence that the hippocampus is involved in a propose-but-verify mechanism (i.e., rapid pattern-separation processes) across multiple exposures during cross-situational word learning. Several neuroimaging studies have also demonstrated that the hippocampus plays a critical role in forming and reconstructing the relational memory representations underlying flexible cognition and social behavior (Montagrin, Saiote, & Schiller, 2018; Rubin, Watson, Duff, & Cohen, 2014). In the process of inferring the meaning of a new word across multiple social situations, successful learners may efficiently verify meanings across a series of scenes, and update and reconstruct information in the hippocampus during the encoding of words.

Although most current models of word learning posit that the hippocampal/medial-temporal systems support initial rapid learning, while the neocortex system represents more gradual, long-term, lexical knowledge over a period of weeks to years (Davis & Gaskell, 2009; Davis et al., 2009), there is recent research on memory that neocortical areas also support rapid memory formation at the time of encoding along with the hippocampus within hours or days, which together contribute to later long-term memory formation (Brodt et al., 2016; Coutanche & Thompson-Schill, 2014; Hebscher et al., 2019; Hofstetter, Friedmann, & Assaf, 2016). Brodt et al. (2016) found that the posterior parietal cortex encodes memories for spatial location during initial encoding of unknown virtual environments, and initial parietal activity predicted later memory performance. Furthermore, rapid cortical learning may be promoted by more elaborate encoding conditions such as involving action selection, active discovery, multimodal learning, and social/emotional processing (Hebscher et al., 2019; Macedonia et al., 2019; Schott et al., 2013). Our findings are consistent with recent perspectives. In the current study, we found that the strength of activation in the initial stage of encoding predicted acquisition of successful semantic knowledge after overnight consolidation (Davis et al., 2009). Successful learners in the social learning condition were more likely to extract relevant social, perceptual- and action-related semantic information, resulting in the rapid formation of memory in relevant cortical areas and the hippocampus, ready for long-term retention.

It is worth discussing the potential long-term effects of L2 social learning, along with findings from previous studies on bilinguals on neuroplasticity (DeLuca, Rothman, Bialystok & Pliatsikas, 2019; Hosoda, Tanaka, Narai, Honda, & Hanakawa, 2013; Mårtensson et al., 2012; Li, Legault, & Litovsky, 2014; Pliatsikas, 2020). These studies have generally reported long-term structural brain changes in the hippocampus and in other cortical and subcortical areas outside of the typical language network, when bilinguals receive intensive training or exposure to L2. Mårtensson et al. (2012) reported that interpreters showed increased cortical thickness in the right hippocampus and the left STG through intensive interpreting training. It was also found that these changes were correlated with individual proficiency levels. Similarly, Hosoda et al. (2013) examined gray-matter density and white-matter integrity among Japanese learners of English after a 16-week intensive vocabulary training on pronunciation and other contextual examples of vocabularies. Hosoda et al. found that L2 vocabulary size and competence were positively correlated with volumetric increase mainly in the right hemisphere, IFG, and STS/SMG. They also observed strengthened connectivity between these areas and subcortical ones. Further studies are needed to understand the precise role of cortical areas and the hippocampus in social L2 learning, and their independent and joint contributions to long-term representation depending on learners’ experience during L2 learning.

5. Limitations, future studies, and conclusion

We conclude with the following three major suggestions for future research: First, the factors or variables playing significant roles in learning from social contexts need to be examined. Although individual differences were observed at the neural level during social learning, the causes of such individual variations are not known. Among many other factors, one cause could be cognitive factors such as working memory and inductive learning (Wen, Skehan, Bediron, Li, & Sparks, 2019). Affective factors such as motivation, emotion, and willingness to communicate may also play a mediating role in learning from social contexts (Dörnyei & Ryan, 2015). Second, this study should be replicated with more learning materials for generalizability. In this study, we used limited types of vocabularies such as verbs, adjectives, and greeting words. In future studies, different types of words, such as nouns and abstract words, and categories of language, such as grammar and pragmatics, need to be investigated. Third, the effects of different types of social learning, for example real interaction and virtual reality, and their long-term impacts on L2 learning (Li & Jeong, 2020) need to be examined. In this study, we used only stimulated videos, although we tried to make them as natural as possible to reflect real interactions.

Despite these limitations, this study provides evidence that the brain mechanisms of L2 learning from social contexts differ significantly from those of L2 learning through traditional classroom-based translation methods. While the former involves the integration of both verbal and nonverbal information, the latter relies on automatic memorization. The latter has also served as the empirical basis of many previous studies of bilingual memory. Herein, we argued that learners who acquire the brain network involved in social, action perception, and memory-related areas at the early stage of learning will be able to acquire and retain L2 knowledge more efficiently. On the one hand, social learning provides better learning opportunities for L2 learners to establish rich semantic representations in the brain, while on the other, serves as an empirical basis for new theories and models of bilingualism and second language acquisition.

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.

Acknowledgments

HJ was supported by the JSPS KAKENHI Grant of Japan (#18K00776). PL was supported in part by a startup grant from the
Hong Kong Polytechnic University. We express our special gratitude to several former colleagues of the Kawashima-lab at Tohoku University, especially Dr. Hiroshi Hashizume and Dr. Satoru Yokoyama, for their support during the MRI data acquisition.

References

Badre, D., & Wagner, A. D. (2002). Semantic, retrieval, mnemonic control and prefrontal cortex. Behavioral and Cognitive Neuroscience Reviews, 1, 206–218.

Barsalou, L. W. (2008). Grounding symbolic operations in the brain’s modal systems. In G. R. Semin, & E. R. Smith (Eds.), Embodied grounding: Social, cognitive, affective, and neuroscientific approaches (pp. 9–42). Cambridge: Cambridge University Press.

Beauchamp, M. S. (2015). The social mysteries of the superior temporal sulcus. Trends in Cognitive Sciences, 19(9), 489–490. https://doi.org/10.1016/j.tics.2015.07.002

Bedny, M., Dravidia, S., & Saxe, R. (2013). Shindigs, braches, and rodeos: The neural basis of event words. Cognitive, Affective, & Behavioral Neuroscience, 14, 891–901.

Beren, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by proposer-but-verify hypothesis testing. Current Biology, 28(7), 1132–1136.

Binney, R. J., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. Neuroscience & Biobehavioral Reviews, 112, 28–38.

Bloom, P. (2000). How children learn the meanings of words. Cambridge, MA: MIT Press.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wobler, T., et al. (2009). Hippocampus activity differentiates good from poor learners of a novel lexicon. NeuroImage, 25, 958–968.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the marshall toolbox for spm 99. NeuroImage, 16, 5479–5485.

Brodie, S., Frith, C., Flanagin, V. L., Gais, S., Glauser, S., Goutoueln, D., & Scholz, M. (2016). Rapid and independent memory formation in the parietal cortex. Proceedings of the National Academy of Sciences, 113(46), 12351–12356. https://doi.org/10.1073/pnas.1607911113.

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of face, social space, and social perception and cognition in the superior temporal sulcus. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 370(1675), 20140009.

Davey, J., Thompsona, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., & Wilson, H. (2013). A neural interactive location for multilingual talent. Journal of Cognitive Neuroscience, 25(1), 103–117. https://doi.org/10.1162/jocn.2012.21171

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as the simultaneous use of two or more modal systems. In A. C. Marslen-Wilson, & M. S. Gaskell (Eds.), Multilingualism (pp. 1–17). Routledge.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences, 361(1536), 3711–3719. https://doi.org/10.1098/rspb.2005.2743.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.

Doherty, J. P., Ashburner, J., Friston, K. J., & Frackowiak, R. S. (2004). Statistical parametric mapping: The analysis of functional brain images. Academic Press.
Schott, B. H., Wüstenberg, T., Wimber, M., Fenker, D. B., Zierhut, K. C., Seidenbecher, C. I., et al. (2013). The relationship between level of processing and hippocampal-cortical functional connectivity during episodic memory formation in humans. *Human Brain Mapping, 34*(2), 407–424.

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage, 84*, 265–278.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Yang, J., Gates, K., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics, 33*, 29–49.

van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S. A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience, 24*(11), 2237–2247.

Verga, L., & Kotz, S. A. (2019). Spatial attention underpins social word learning in the right fronto-parietal network. *Neureomage, 195*, 165–172.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. E. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia, 40*(12), 2136–2143.

Vong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping, 28*(10), 995–1006.

Wen, Z. E., Skehan, P., Biedron, A., Li, S., & Sparks, R. L. (2019). *Language aptitude: Advancing theory, testing, research and practice*. Routledge.

Zhang, X., Yang, J., Wang, R., & Li, P. (2020). A neuroimaging study of semantic representation in first and second languages. *Language, Cognition and Neuroscience, 1–16*. https://doi.org/10.1080/23273798.2020.1738509.