Initial signs of learning: Decoding newly-learned vocabulary from neural patterns in novice sign language learners

Megan E. Hillis* & David J. M. Kraemer

Department of Psychological and Brain Sciences, Dartmouth College

*Corresponding author: Megan E. Hillis (megan.e.hillis.gr@dartmouth.edu)

Keywords: cognitive neuroscience, sign language, language learning, fMRI, semantic representation

Abstract

How do novice language learners represent semantic information in their new language? The extent to which multiple languages are supported by divergent or overlapping semantic representations in bilinguals has been well-studied, but less is known about how new knowledge is integrated into established representational networks at the earliest stages of acquisition. Furthermore, examining language across modality (sign vs. speech) can provide unique insight into language unconfounded by perceptual features. We present two experiments in which hearing non-signers underwent brief training in American Sign Language (ASL) followed by fMRI scanning. Across both datasets (N=50), we use representational similarity analysis (RSA) to identify brain regions where neural patterns reflect semantic relationships between stimuli. In Study 2 (N=40) we show that multivariate neural measures of semantic representation in several frontal, temporal, and occipital regions reflect individual participant-level comprehension. These results demonstrate the role of frontal and temporal regions, especially bilateral superior temporal sulcus, in representing semantic content across language and modality in novice learners.

Introduction

How do novice language learners represent semantic information in their new language? The extent to which multiple languages are supported by divergent or overlapping semantic representations in bilinguals has been the topic of much research, but less is known about how new knowledge is integrated into established representational networks, especially in adults at the earliest stages of learning. Furthermore, most investigations have compared languages which share a modality (i.e., spoken languages). Signed languages can provide a unique lens into language unconfounded by perceptual qualities that spoken languages share. Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) have proven powerful tools to assess the overlap and divergence across languages in the brains of both fluent multilinguals and novices who are attempting to acquire a new language. In particular, multivariate pattern analysis methods can leverage the brain's remarkable ability to represent conceptual knowledge in a way that is at least partially amodal[1,2] to characterize how and when representations are shared across languages, and whether the emergence of this overlap can track individual participant-level comprehension.

Research in bilinguals has found evidence for both commonality and divergence between neural processes underlying multiple languages in the brain, and it appears that for non-native languages these complex interactions are sensitive to age of acquisition and proficiency. A series of meta-analyses have demonstrated that in bilinguals, languages rely on similar networks in the brain but may utilize them in subtly different ways[3,4] Semantic processing may rely on overlapping representations which are at least partially language-independent, while perceptual processing is unsurprisingly associated with language-specific features. In one example, Honey and colleagues[5] found that in English-Russian bilinguals who listened to the same story in each language, neural responses were language-specific in early auditory areas, but activations in anterior temporal, parietal, and frontal cortices were shared across languages.

Concerning semantic processing in particular, multivariate pattern analysis methods such as Representational Similarity Analysis (RSA)[6] have provided a way to probe the correspondence of fine-grained patterns of activity with an a priori model, such as a categorical structure or patterns of semantic features. This body of work generally asserts that semantic information is "amodal" if it is sensitive to conceptual properties of stimuli (e.g. object category) without influence of the modality of the stimulus cue. It is well-established that such methods can decode conceptual representations that are shared between words and pictures[1,2,7,8] as well as written and audio presentations of the same content[9–12]. This property also extends to homologous words across languages. Cross-language decoding techniques have revealed that representations of fine-grained distinctions between nouns are language-independent in left temporal regions[13,14], that certain tasks can induce cross-language decodability in sensory cortices (e.g. occipital regions for word reading, pre-and postcentral regions for production[15], and

that cross-language representations may be more strongly evoked by deeper semantic processing[16].

These studies demonstrate that bilinguals access language in a way that is at least partially modality-independent, especially where semantic processing is concerned. However, less is known about how these partially overlapping representations develop as individuals learn a new language. Some studies have found that late and less proficient bilinguals exhibit more overlap across languages than native bilinguals[17,18], which could be attributed to use of established native language networks to scaffold new learning. Leonard and colleagues[19] demonstrated using MEG that in unbalanced bilinguals, responses to the non-dominant language were localized to similar networks but occurred more slowly, and recruited more right hemisphere and posterior visual areas which may be less necessary for more proficient users. This shift over time from reliance on perceptual areas to greater involvement of semantic-conceptual regions has been observed over relatively short (months) periods of learning.

Only one study to our knowledge has utilized this type of cross-decoding in *bimodal* bilinguals – those fluent in two languages that differ in modality (speech versus sign). Evans and colleagues[20] presented bilinguals of spoken English and British Sign Language (BSL) with clips of each language during fMRI scanning. Using RSA[21], they identified areas where neural activation patterns reflected both item-level semantic relationships and broader category-level relationships between the stimuli in each language. They found evidence of modality-specific (within-language) semantic representation at the item level, as well as overlapping cross-language representations of the broad categories. The latter is consistent with a large body of research including lesion studies and functional neuroimaging that suggests that, overwhelmingly, neural mechanisms underlying linguistic processing of sign and speech are very similar and rely on a left-lateralized perisylvian network (for review, see MacSweeney and colleagues[22]).

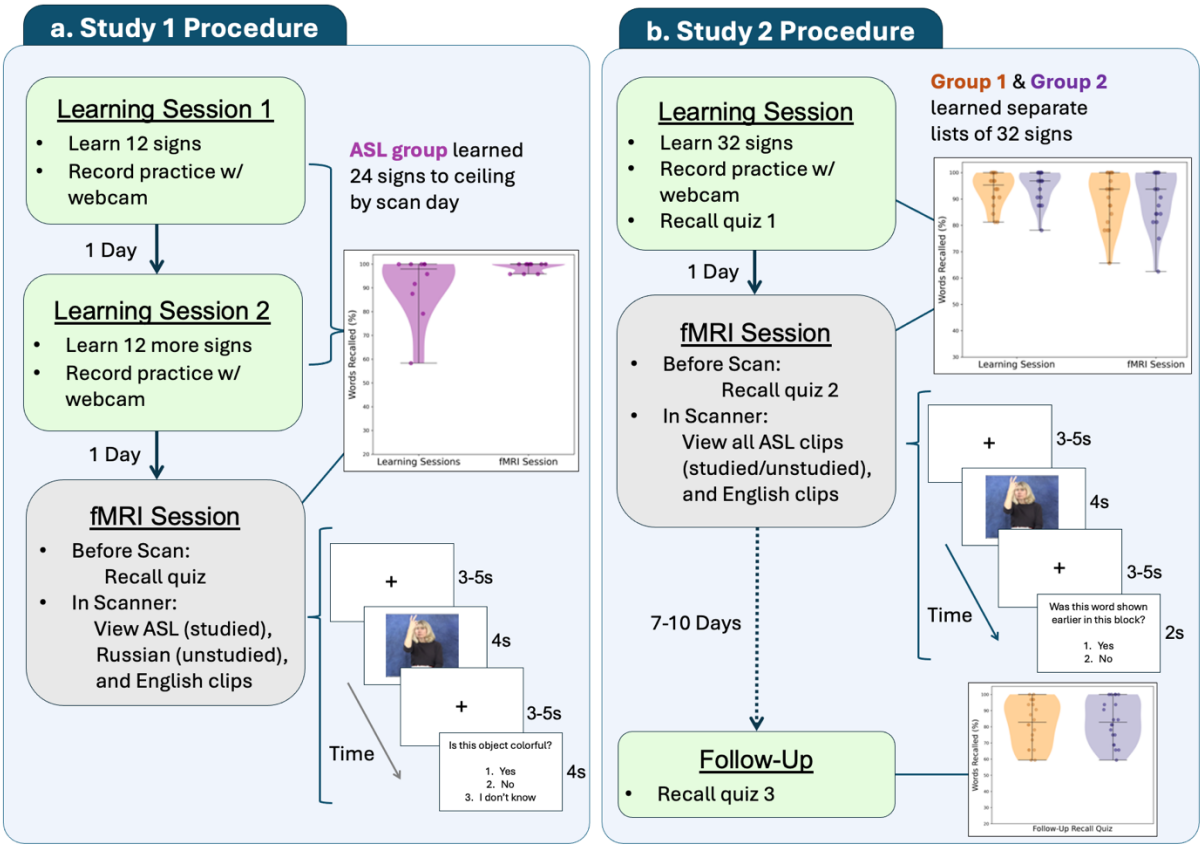
While it is clear that in native signers, left inferior frontal[23–28] and superior temporal[29–32] areas serve linguistic functions regardless of modality, no previous studies have attempted to use this type of method to characterize semantic representation in the early stages of sign language learning. Univariate investigations in this area have found that compared to native signers, hearing late signers make less use of right parietal areas implicated in spatial processing of sign[33–35], and show neural patterns that overlap with those evoked by their native spoken language[19,36], similarly to novice learners of spoken languages who may use their established native language as a scaffold for subsequent learning[17,37]. Neural overlap between sign and speech also appears to increase with proficiency in the early stages of sign language learning. For example, an fMRI study by Williams, Darcy, and Newman[38] found that in participants scanned repeatedly over the course of 10 months of ASL training, lexico-semantic processing of ASL in the left inferior frontal gyrus (IFG), supramarginal gyrus (SMG), and putamen increased at each timepoint.

However, the inferences that can be drawn from similarities in univariate activation are limited. Multivariate pattern analysis techniques such as RSA, whereby the relationships between stimuli are modeled in a high-dimensional space and the presence of neural activity with a similar pattern of relationships is assessed, can provide uniquely fine-grained insights into the content of linguistic processing. The present research leverages methods pioneered in studies of concept knowledge[39,40], especially in physics and engineering[41–43] to evaluate the correspondence between individual-level neural "scores" and knowledge measured by traditional pencil-and-paper tests.

We present two studies in which hearing spoken language users underwent brief (2 or 3 half-hour sessions) training in ASL, then participated in fMRI scanning, during which neural activity was recorded while they watched clips of well-known (English), newly studied (ASL), and unstudied linguistic content. An overview of both training procedures as well as recall performance for the studied content is illustrated in Figure 1. In Study 1, participants were trained to ceiling on a set of 24 ASL nouns, to test the ability of these analysis techniques to distinguish known from unknown semantic content on the group level when comprehension was high. A previous preliminary analysis of these data demonstrated that in areas that reflected semantic relationships between the studied nouns in English, a classifier trained on ASL trials could classify held out items into broad semantic categories (animals, fruits, vehicles) significantly above chance[44]. In Study 2, the vocabulary list was expanded and the training time shortened to produce individual differences in recall performance. Among these participants, we demonstrate that the extent of decodability of these item-level representations tracks with individual behavioral measures of comprehension. These results demonstrate that neural patterns associated with emerging knowledge in novice signers are largely consistent with existing studies of spoken language learners, and that novice signers exhibit overlapping representations of their two languages on the fine-grained item level. Indeed, we also demonstrate that the *degree* of this correspondence in several regions is meaningfully related to how accurately individuals could translate the signs to English, underscoring the role of frontal and temporal regions, especially bilateral superior temporal sulcus, in representing semantic content across language and modality in novice learners.

Figure 1

Learning procedure and quiz scores for studies 1 and 2



Note. Procedure for the two studies. Study 1 participants were trained to ceiling on a set of 24 ASL signs over two online lessons and one online review. Behavioral performance at the end of the online training and immediately before the scan are shown in panel a. Individual quiz scores are shown as a scatter plot overlaid on the violins. The horizontal lines indicate the maximum, mean, and minimum score at each timepoint, while the curves represent the frequency of the result in the distribution. Study 2 decreased the length of training and increased the amount of content to 32 signs. Additionally, participants completed a third recall quiz approximately one week after the fMRI session, providing a measure of how well the content was retained over a longer period, shown in Panel b. Screenshots of the ASL stimulus videos shown in the fMRI procedure schematic are credited to the ASL-LEX project (https://asl-lex.org/), used with permission.

Study 1 Methods

Participants

Twenty-two volunteers from the local community participated in this study. Data from two participants were excluded due to incomplete scans, resulting in a sample of N = 20 (13 female, mean age = 20 years, SD = 1.70). All participants were fluent English speakers who reported no prior knowledge of American Sign Language or Russian. Participants provided informed consent prior to participation in each day of data collection and were compensated either with curricular extra credit points or a gift card. All protocols were approved by the Dartmouth Committee for the Protection of Human Subjects. (https://www.dartmouth.edu/cphs/)

Stimuli and Design

Stimuli for the behavioral and scanner tasks (i.e. the language lessons and semantic task, detailed in the section "Procedure") were short audiovisual clips each containing a single noun in ASL, Russian, or English. The ASL videos were provided by ASL-LEX, a database of lexical and phonological properties of ASL signs[45,46]. The Russian and English videos were created with efforts to mimic the style of the ASL-LEX videos by lab volunteers who are fluent in the respective languages. Each video clip consisted of the presenter seated before a neutral background, demonstrating a single vocabulary word or sign. Of the 24 stimuli, 12 were members of the semantic categories of interest (animals, fruits, and vehicles), while another 12 were selected as distractor items to obscure the intended categories from participants. To ensure that participants would not be able to guess the meanings of the clips without training, ASL signs which were rated greater than average in transparency (the extent to which a sign's meaning is knowable by non-signers due to iconicity, or form that resembles meaning) by a sample of hearing non-signers collected by the creators of the ASL-LEX corpus were excluded from the stimulus set. Additionally, all ASL and Russian stimuli were pilot tested on Amazon Mechanical Turk to confirm that native English speakers with no training in ASL or Russian were not able to guess the meanings of the stimuli and that subjective ratings of visual similarity (for ASL) and auditory similarity (for Russian) between them did not correlate with object category.

Semantic and Perceptual Models

Word2Vec. To model semantic relationships between the stimuli, we calculated the representational distance between each pair of items using a word2vec model[47]. Cosine dissimilarity between each pair of target items was calculated with pretrained embeddings derived from the full corpus of English wikipedia[48]. The resulting dissimilarity matrices are used as an a priori model of semantic relationships between the items.

Phonology. A potential confound in ASL is that semantic similarity between two or more signs is sometimes correlated with visual similarity of the signs. Broadly, there are two ways for

this similarity to emerge in ASL: phonology (i.e., the low-level features of hand movement, finger position, palm orientation, etc. which characterize a given sign), and iconicity/transparency (discussed in the following section). To model perceptual features of the stimuli which may not be entirely orthogonal to semantics (i.e., phonology), we constructed models of phonological properties for each stimulus list. For the ASL stimuli, we modeled phonological properties of individual signs using the metadata codes provided by the ASL-LEX 2.0 database[46]. These codes were created based on the Prosodic Model[49] to describe the structural composition of signs, encompassing handshape, palm orientation, location in sign space, movement, and sign type (one- or two-handed, symmetry) as well as other features. We calculated phonological similarity between two signs as the number of phonological features that each pair of signs shared.

Iconicity/Transparency. Iconicity, or the non-arbitrary relationship between linguistic form and meaning, is a property found in many languages which is particularly pervasive in signed languages. It may be of particular importance to the study of novice signers, considering that iconic signs are remembered more accurately[50–52] and are recognized with faster response times[53,54] during the early stages of sign language acquisition. Meanwhile, in fluent signers, the behavioral effects of iconicity are weaker and highly task dependent[55]. The ASL-LEX 2.0 database provides mean subjective iconicity ratings by a group of hearing non-signers, which we adopt as our non-signer iconicity model. For one sign ("pineapple"), non-signer iconicity ratings were not available, so for this single case we substituted the reported mean transparency (here, defined as the nonsigners' subjective ratings of how obvious the sign's meaning would be to others). Iconicity is distinct from transparency, which refers to the ease with which the meaning of a sign can be guessed by its form alone, but in our sample of very novice signers these constructs likely capture similar variance.

The dissimilarities between the stimuli estimated by each model were first z-scored and rescaled from zero to one, where zero indicates the exact same item and one indicates the maximal distance. The patterns of dissimilarity between the full set of target items in all three of these models, along with their Spearman correlation with each other, are presented in Figure 5. Before correlation with the ASL neural DMs, all models within each language were additionally orthogonalized to each other using linear regression.

Procedure

Participants completed three short (approx. 30 min) online behavioral sessions on the two days preceding and the day of the fMRI scan. Half of the participants (N=10) were assigned to learn the set of 24 concrete nouns in ASL, while the other half learned the same nouns in Russian (although the Russian group is not discussed in the present paper due to poor learning performance). The lessons were administered through Qualtrics (Qualtrics, Provo, UT). During each of the first two learning sessions, participants learned 12 nouns in their target language

through watching and mimicking the expert videos. They then completed a set of multiple-choice questions, a free recall task, and finally used their computer's webcams to record videos of themselves practicing each sign. In the third practice session, which occurred on the same day as the fMRI scan, participants reviewed all 24 nouns that they had previously learned, created a final set of webcam recordings of themselves performing each sign, and completed another free recall quiz before arriving for the fMRI scan session. During the fMRI session, participants watched the same audiovisual clips followed by questions which probed either the semantic meaning of each noun (such as "Is this object colorful?" or "Would it be easy to cause this object to move?") or non-semantic perceptual features of the clip ("Has this clip been presented already in this block?"). They answered this question by pressing a button with their right index finger or middle finger. All participants, regardless of which language they had studied in the learning period, viewed clips in ASL, Russian, and English. ASL and Russian were presented in counterbalanced blocks of 16 trials each during the first two functional runs. The English stimuli were presented in similar blocks of 16 trials during the third functional run, due to concern that knowledge that the same 24 nouns were presented in each language might help participants "guess" the meanings of clips in the unstudied language. The non-semantic question trials were included to encourage participants to pay attention even during blocks when they did not know the semantic meanings of the clips. Each clip was presented twice in each language.

fMRI Data Acquisition

Brain images were acquired using a 3 Tesla Siemens PRISMA fMRI scanner with a 32-channel head coil. A single high-resolution T1-weighted anatomical scan and three 8-minute functional runs were performed for each participant. Each 2D EPI sequence consisted of 192 measurements with a 240 mm2 field of view to provide full brain coverage over 46 slices (Flip angle = 79°; TE = 32 ms; TR = 2500 ms; 3x3mm voxels). In the scanner, stimuli were presented using PsychoPy[56] version 2021.2.3 (using Python 3.6).

Image Preprocessing and General Linear Model

Brain images from both datasets were preprocessed with fMRIprep version 21.0.1[57,58], which is based on *Nipype* 1.6.1[59]. As recommended by the software developers, boilerplate text generated by fMRIprep is reproduced exactly in the Supplementary Materials. In summary, preprocessing included skull stripping, slice-time correction, smoothing, registration and normalization. fMRIprep also estimates the time course of confounds such as head motion, which are discussed below.

Beta Estimation. After preprocessing, a univariate regression model using the GLM was then calculated at the trial level, such that beta-value estimates for each stimulus and their temporal derivatives were generated separately for each run using Nilearn's FirstLevelModel tool[60]. Additional regressors of no interest were also included in the model: the time course of six motion parameters (translation and rotation in the X, Y, and Z dimensions) generated by

fMRIprep, and spike regressors for any timepoints with framewise displacement greater than three standard deviations above the mean. Brain activity was sampled from the initial presentation of the video stimulus through a short intra-trial fixation and a 4s semantic response period. Finally, for stimuli which appeared in more than one run, beta estimates were additionally combined across runs with an item-level fixed effects model, yielding a single contrast estimate for each item in each language.

Dissimilarity Matrices. To assess the correspondence between neural activity patterns and the hypothesized relationships between stimuli, we constructed neural dissimilarity matrices (DMs) of the correlation distances between the neural responses to each item in each language. One DM was constructed for each subject at each parcel in the Schaefer cortical parcellation atlas[61] (500-parcel version).

Representational Similarity Analysis

After computing the item-level neural dissimilarities in each parcel as described above, we used RSA to assess the correlation between the neural data and a priori models for each language. As diagrammed in Figure 3, the English neural data was correlated with the word2vec model. Separately, the ASL and Russian neural data were correlated with a word2vec model that was orthogonalized to one or more additional models, phonological and (for ASL only) non-signer rated iconicity, as described under Stimuli and Design. This was done to remove any correlation that could have been attributed to non-semantic features of the stimuli. Spearman correlation between the parcel-level DM and the model was passed through a Fisher z-transformation and compared to a null distribution calculated as the dot product of the true model and a randomly permuted model standardized over 1,000 iterations. The resulting correlations were subjected to a one-sample t-test at each parcel to identify brain regions where the neural correlation with the semantic model was greater than zero, in other words, areas where BOLD activity patterns reflected semantic relationships between the items.

Study 1 Results

Recall Quiz Performance

At the end of each training session as well as the review session prior to the fMRI scan, participants completed a free recall quiz during which they were shown one of the clips from the training material and asked to type its English translation. Due to poor recall performance (M_{lesson}=52.9%, SD_{lesson}=6.0%, M_{scan}=62.9%, SD_{scan}=5.9%), the Russian group was not analyzed further. The distributions of recall scores at each timepoint for the ASL group are shown in Figure 1 (M_{lesson}=91.3%, SD_{lesson}=4.2%, M_{scan}=98.7%, SD_{scan}=0.6%). These results indicate that the ASL group learned these 24 signs to ceiling by the final day. Thus, these participants provide

a useful lens by which to measure representational overlap between a newly learned language (ASL) and a well-known language (English) in novices who have successfully learned a small set of signs.

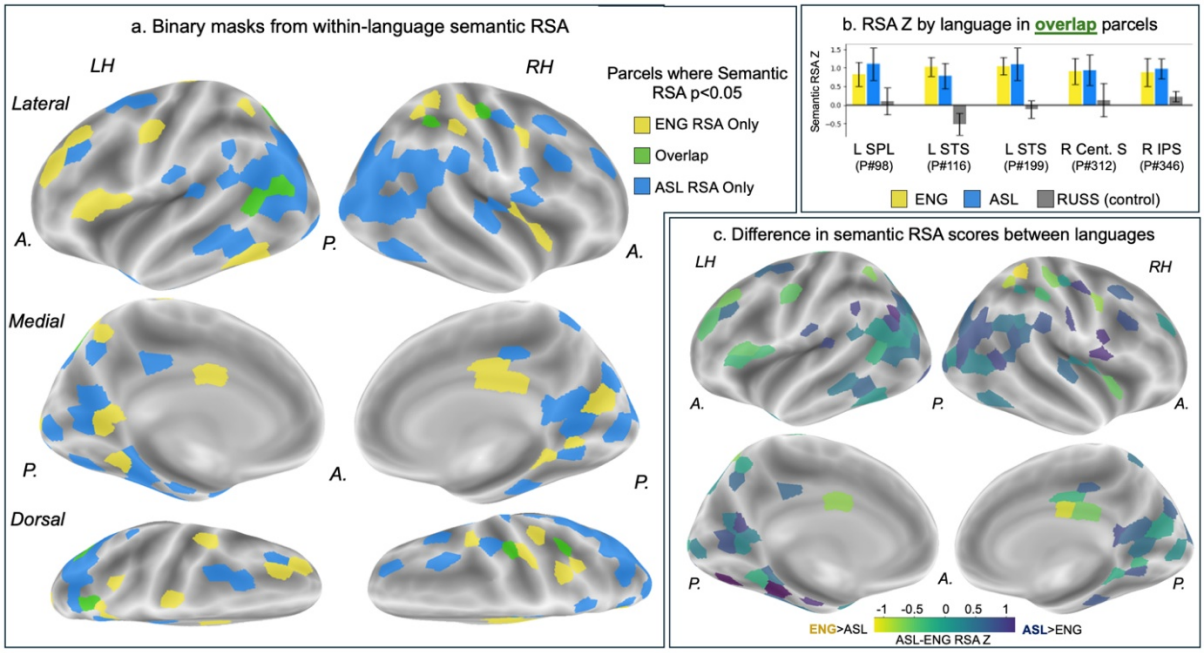
Representational Similarity Analysis

Parcels where correlation was greater than zero at the $p<0.05$ level for ASL and English in the ASL group are shown in the overlap map in Figure 2. A total of 76 neural parcels encompassing areas including bilateral parietal and occipital areas were correlated with word2vec for ASL. Meanwhile, a total of 34 parcels including left inferior frontal and bilateral parietal areas were correlated with the semantic model in English. Of these, five parcels (displayed in green) were significantly correlated with word2vec in both languages, located in the left superior parietal lobule (parcel 98), left superior temporal sulcus (parcels 117 and 199), right central sulcus (parcel 312), and right intraparietal sulcus (parcel 346).

As a negative control for novel language stimuli that the participants had *not* learned, we also computed RSA correlation between the neural data from the Russian trials and the word2vec model, as well as a phonological model (constructed from pairwise auditory similarity ratings made by non-Russian speakers). The resulting RSA map is shown in Supplementary Figure 1A. Notably, no parcels in the left language network correlated with the semantic model in the unstudied language.

Figure 2

Study 1 semantic RSA results



Note. Locations of significant parcels from the semantic (word2vec) RSA for each language in Study 1 are shown on the FSAverage6[62] semi-inflated brain surface. (a.) Parcels where the permutation-corrected Z values differed from zero with $p < 0.05$ are shown, with parcels which were significant in both languages highlighted in green. (b.) Semantic RSA scores for the five overlap parcels in each of the three languages. Bars represent standard error. Notably, none of these areas identified as overlapping between ASL and English showed significant fit of the word2vec model to the Russian stimuli, which were unknown to the participants. (c.) Difference in semantic RSA scores between ASL and English (greater=ASL>ENG, lesser=ENG>ASL, 0=languages were equally correlated with semantic model).

Interim Discussion

These results provide a proof of concept that even after very brief training, both shared and language-dependent semantic representations of stimulus meaning can be decoded from neural activity. This is consistent with the findings of Evans and colleagues[20] that representations of sign and speech are partially overlapping. Study 1 participants were trained to ceiling on the target ASL signs to give the group analysis the best chance possible to identify brain regions that correlate with the word2vec model in novice learners. However, because all participants performed similarly, we cannot assess from this study alone whether the *degree* of correlation with the semantic model predicts how well participants have learned the material. To this end,

we conducted a second, larger study, which followed a similar paradigm with efforts to induce more variation in performance between participants.

Study 2 Methods

Participants

Forty-three participants who were fluent English speakers with no prior training in ASL participated in this study. Three were excluded (one due to failing to complete the pre-scan quiz, and two due to excessive motion in the scanner) for a final N = 40 (28 female, 2 nonbinary, mean age=20.3 years, SD=3.0). Participants provided informed consent at the beginning of each session and were compensated either with curricular extra credit points or a small monetary reward. All protocols were approved by the Dartmouth Committee for the Protection of Human Subjects (https://www.dartmouth.edu/cphs/).

Stimuli and Design

In this study, participants were divided into two groups of ASL learners who were each assigned to learn a different vocabulary list. Thus, for each group, there was a set of 32 studied signs, and a separate set of 32 signs which were never shown during the training phase. As before, each stimulus list contained four items from each of the target categories (animals, fruits, vehicles), and the rest were distractor items. ASL sign clips were selected from the ASL-LEX database with the same procedure described in Study 1 methods. Audiovisual clips of homologous English words were created by a lab volunteer with efforts to mimic the style of the ASL-LEX videos. As before, the transparency of the ASL clips was pilot tested on Amazon Mechanical Turk to confirm that English speakers with no training in ASL could not guess the meanings of the signs from their visual features alone.

Procedure

Study 2 followed a similar training paradigm as the previous study, whereby participants completed brief activities online before an fMRI scan where they were presented with clips of studied and unstudied signs. However, while previously we optimized our pre-scan training procedure to ensure participants learned the ASL vocabulary to ceiling, here we decreased the length of training (from two hours beginning two days prior to the scan to just 40 minutes beginning one day before) and increased the number of signs (24 to 32) to increase the difficulty and ensure variation in performance across the group. The vocabulary lessons were administered through Qualtrics (Qualtrics, Provo UT), and guided participants first to watch and mimic the sign videos, complete brief multiple-choice assessments, use their devices' webcams to record themselves practicing each sign, and finally complete a recall test where they were prompted to type in the English translation for each sign video. Recall quiz 1 scores were calculated as the proportion of items the participant correctly translated.

On the day of the fMRI session, participants took another brief recall quiz before the scan. All 64 signs were presented in an old-new paradigm, where participants were asked to indicate whether they had studied each sign, and for each sign marked "old", they were asked to type in the English translation. The proportion of studied signs for which they provided the correct translation was calculated as their pre-scan (quiz 2) recall score. In the scanner, after the initial anatomical scan, clips from both the studied and unstudied lists were presented in a randomized order in three functional runs. Each clip was shown twice per run, resulting in six presentations per video. On a small number of trials (10%), the video clip was followed by a non-semantic probe ("Have you seen this clip already during this run?") and participants answered yes or no by pressing a button with their right index or middle finger. The semantic task from the previous study was omitted here to remove any possibility that the questions were cuing semantic features (such as color) that were confounded with object category. After three functional runs of ASL clips, participants repeated this same task with the audiovisual English clips. As before, English was presented at the end to prevent participants from knowing the full set of targets and increasing their odds of "guessing" the meanings of the unknown signs.

One week after the fMRI session, participants were contacted by email and asked to take the same old-new recall test a final time (quiz 3) to assess longer-term retention of the signs. The complete timeline of the three time points and quiz performance at each are diagrammed in Figure 1.

fMRI Data Acquisition and Preprocessing

Brain images were acquired using the same equipment and measurement parameters as Study 1; only the length and number of runs varied. A single high-resolution T1-weighted anatomical scan and six functional runs of 241 measurements (6.3 minutes) were performed for each participant. As before, data were preprocessed with fMRIprep version 21.0.1[57], the boilerplate output of which can be found in the supplementary materials.

Beta Estimation and Dissimilarity Matrices

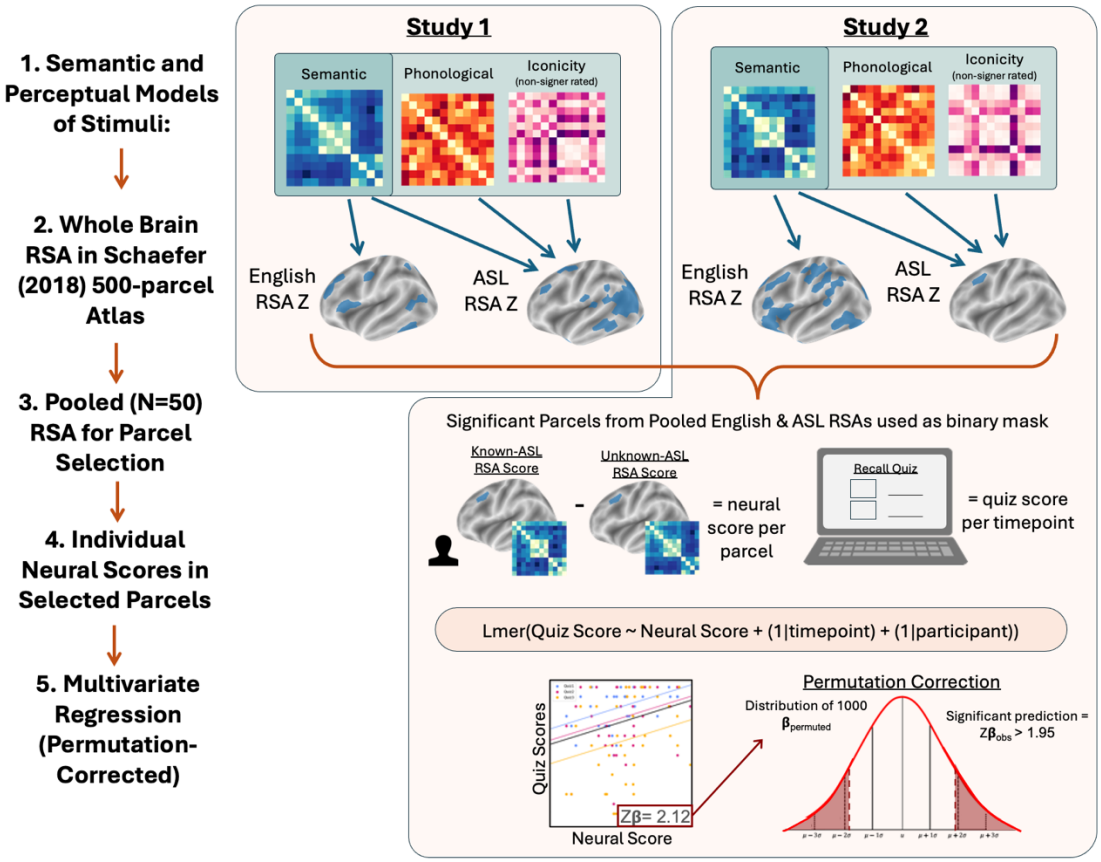
Beta values for each stimulus clip were extracted using the same procedure as Study 1, with one notable difference. For Study 2, because the semantic task was omitted and the N-Back question only followed 10% of trials, brain activity for each was sampled only from the stimulus presentation interval, and the button press interval was modeled as an additional regressor but not analyzed. As before, stimuli which appeared in more than one run were combined with an item-level fixed effects model, yielding a single contrast estimate for each item in each language. Correlation distance between each pair of items was again computed for each cortical parcel in the Schaefer 500-parcel atlas[61], and organized into a single DM for each parcel in each participant.

Pooled RSA for Parcel Selection

As a feature reduction step, we first sought to identify a subset of brain regions which were sensitive to semantic information about the items. Given the evidence that both overlapping and distinct brain regions support sign and speech comprehension, subsequent analyses will consider both parcels which show correlation with the semantic model in English (which all participants spoke fluently) as well as the subset of ASL signs that were known to each participant. To optimize statistical power, we included data from both studies (pooled N=50). This procedure is diagrammed in Figure 3. For each participant, correlation between their neural DM at each parcel and the semantic model derived from word2vec was calculated with RSA. The same permutation testing procedure (whereby each neural DM is compared to a distribution of permuted semantic DMs) was employed, and the resulting correlation values were subjected to a one-sample t-test against zero (no correlation between neural and semantic DMs) for the entire pooled sample. As described in the Study 1 Methods, we orthogonalized the semantic DM for each ASL stimulus list to two additional models, to remove variance associated with visual phonological features and non-signer rated iconicity.

Figure 3

Analysis procedure for studies 1 & 2



Note. Analysis procedure for studies 1 and 2. Initially, group-level correlation with relevant models for each language (semantic only for English, semantic, phonological, and non-signer rated iconicity for ASL) was assessed with RSA. Parcels for which pooled (N=50) RSA Z scores were significant ($p<0.05$) in one or both languages were used as a binary mask for subsequent analyses in Study 2. For Study 2, where participants showed individual differences in recall performance, we computed "neural scores" as the difference between correlation of the known items and unknown items with the semantic model. Then, we fit a linear mixed model at each parcel in the mask to assess whether these scores were predicted by behavioral quiz scores. The observed beta values of these models were contrasted against a distribution of permuted beta values created by shuffling the data and refitting the model 1,000 times at each parcel.

Study 2 Results

Behavioral Performance

Both groups learned the majority of the material, but unlike the previous study, there was individual variation in quiz scores indicating that not everyone had learned all 32 signs to ceiling. Figure 1 shows the target item recall scores for the two groups for all three timepoints (M_{T1}=94.3%, SD_{T1}=1.0%; M_{T2}=90.5%, SD_{T2}=1.6%; M_{T3}=82.7%, SD_{T3}=2.1%). A two-way mixed ANOVA confirmed that the two groups performed comparably at each time point ($F_{group}=0.15_{(38,76)}$, $p=0.70$), therefore we analyze them as a single, pooled sample.

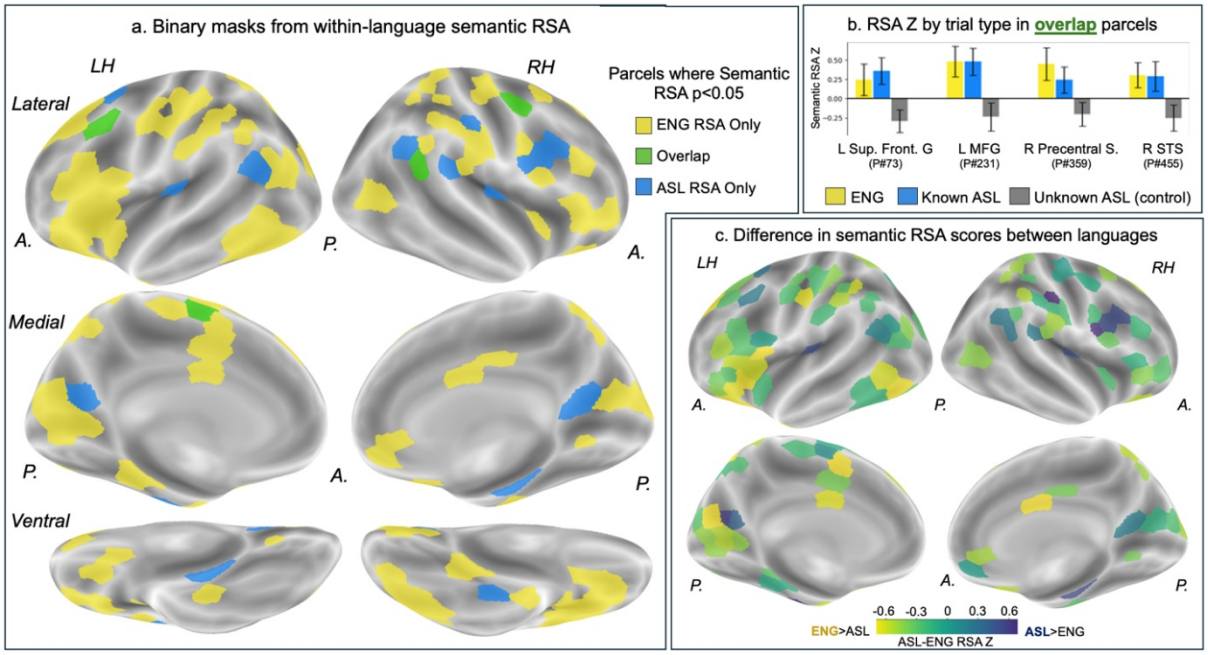
Pooled RSA for Parcel Selection

Considering participants from the ASL group in Study 1 and both ASL groups from Study 2 (pooled N=50), we identified 102 parcels in which the neural responses to English were significantly correlated with word2vec, indicating that they were sensitive to the semantic relationships between the English stimuli. Additionally, 19 parcels showed significant correlation between the known ASL trials and word2vec, four of which overlapped with the English results, located in the left superior and middle frontal gyri (parcels 73 and 231), the right precentral sulcus (parcel 359), and right superior temporal sulcus (parcel 456). The union of the ASL and English results (shown in Figure 4) was used as a binary mask for all subsequent analyses.

Additionally, the RSA results for each of the three models applied to each participant's known ASL stimuli as well as a diagram of the weighting procedure is shown in Figure 5. Neural activity was found to correlate with the phonological and non-signer-rated iconicity models primarily in visual areas, while semantic representation was more distributed across the brain. As a negative control for the relationship between the items in the absence of semantic knowledge, results of the same procedure repeated with each participant's unknown ASL stimuli is shown in Supplementary Figure 1B. Notably, unknown ASL signs appeared to evoke much stronger and more widespread correlation with the phonological model, perhaps because in the absence of knowledge about the signs' semantic meanings, participants attended more to phonological features as a strategy to complete the N-Back task.

Figure 4

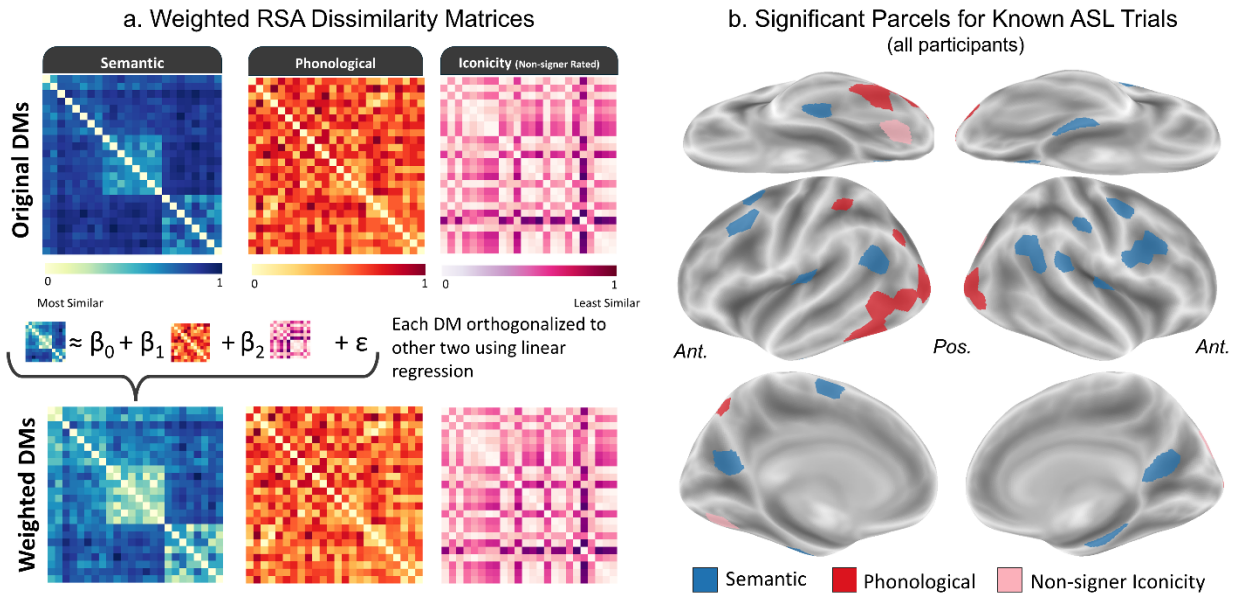
Studies 1 & 2 pooled (N=50) semantic RSA results



Note. Locations of significant parcels from the semantic (word2vec) RSA for each language are shown on the FSAverage6[62] semi-inflated brain surface. Parcels where the permutation-corrected Z values differed from zero with $p < 0.05$ are shown, with parcels which were significant in both languages highlighted in green. (a.) Parcels where the permutation-corrected Z values differed from zero with $p < 0.05$ are shown, with parcels which were significant in both languages highlighted in green. (b.) Semantic RSA scores for the four overlap parcels in each of the three stimulus lists. Bars represent standard error. Notably, none of these areas showed significant fit of the word2vec model to the unstudied list of ASL signs. (c.) Difference in semantic RSA scores between Known ASL and English (greater=ASL>ENG, lesser=ENG>ASL, 0=languages were equally correlated with semantic model).

Figure 5

Weighted RSA: Known ASL trials



Note. Dissimilarity matrices for the three models of ASL stimuli (semantic, phonological, and iconicity) are shown before and after the weighting procedure. Each DM was orthogonalized to the other two using linear regression and rescaled from 0 to 1, resulting in three models that account for only the unique variance of each set of features. On the right, pooled (N=50) RSA results for known ASL stimuli for all three models are shown on the semi-inflated FSAverage6[62] surface. Locations of parcels where permuted RSA Z was greater than zero with $p<0.05$ are shown for each model. No parcels were significantly correlated with more than one model.

Neural Scores Predicting Quiz Performance

In the 117 cortical parcels where neural activity patterns were significantly correlated with the semantic model in one or both languages at the *group* level, we then investigated whether the extent of semantic representation in each *individual's* neural dissimilarities reflected their understanding of the ASL content, as measured by the recall quiz at each timepoint. While the Study 1 participants all performed similarly well on these quizzes, the Study 2 participants displayed enough variation in scores to allow us to explore this question.

If a given brain area is representing semantic content that is only knowable to the subset of participants who studied each stimulus list, we would expect correlation between the neural data and the semantic model to be greater for known signs than unknown signs. For each of the 117 selected parcels, we calculated the delta of the RSA correlation between the neural DMs and

word2vec for the known and unknown sign lists as each participant's "neural score" for that parcel. The calculation of this score is diagrammed in Figure 3c. The relationship between these neural scores and behavioral quiz performance was then assessed using a linear mixed effect model, where neural score was entered as a fixed effect and intercepts for the three time points and for each participant were entered as random effects. These models were created using Pymer4[63], a Python implementation of the R package lme4[64]. To control for false positives, the observed beta values for neural score as a predictor of quiz score were then compared to a null distribution of beta values created by shuffling the mapping of behavioral to neural scores and rerunning the model over 1,000 iterations, simulating the beta values which could occur by chance given the distribution of the data in that parcel[65]. We define significant prediction as an observed beta value greater than 1.95 standard deviations above this permuted null distribution, a critical Z value which corresponds to an alpha level of 0.05 for a two-tailed test.

Five parcels met these criteria, each of which is highlighted in Figure 6. These areas are the left inferior temporal gyrus (parcel 148), the left IFG (parcel 220), the right STS (parcel 456), and the bilateral cuneus (parcels 33 and 281). All five of these parcels show a positive relationship between neural score and quiz score; no parcels negatively predicted quiz score. Notably, the right STS parcel was also one of the four parcels which was significant in both the English and ASL RSA results, indicating that this is a region of semantic overlap between the languages. The observed beta values, Z-scores of those values in the permuted null distribution, and additional features of the model fit at each significant parcel are reported in Table 1.

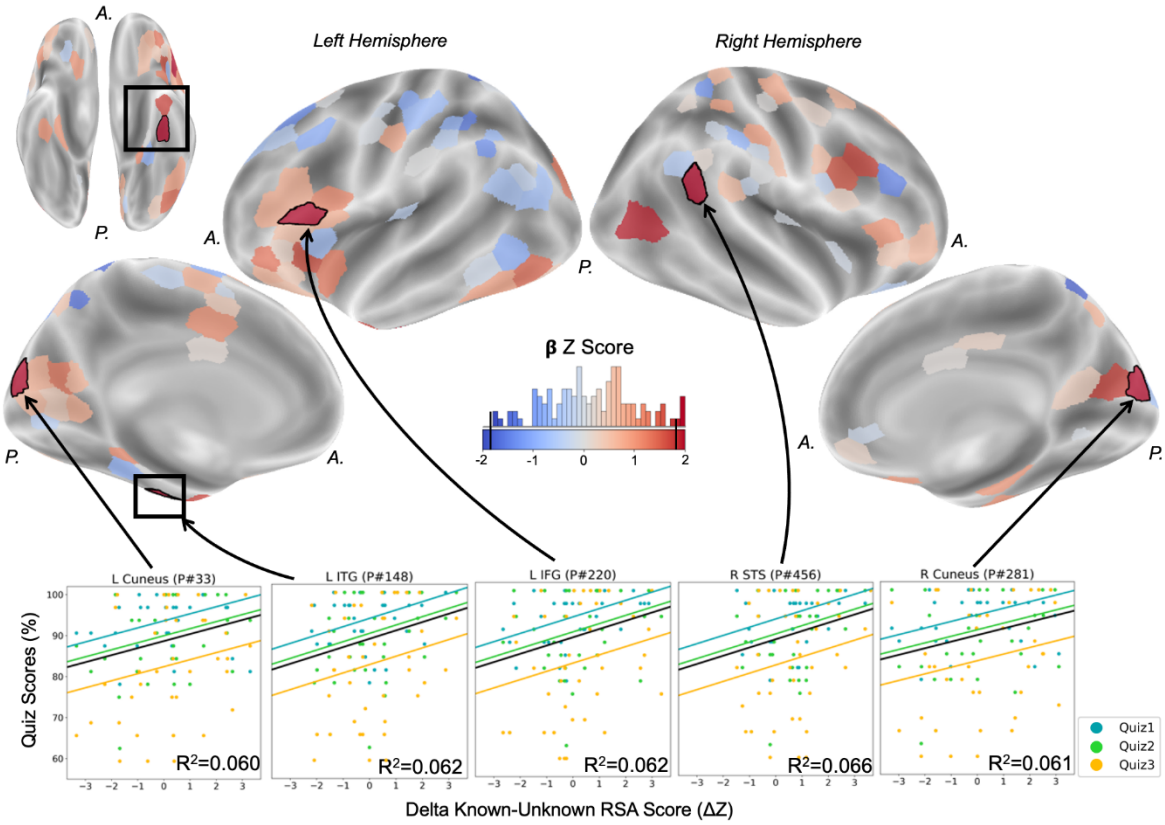
Table 1. Regression results for parcels in which neural scores best tracked quiz score

Parcel	Location	$Z\beta$	β	CI	Marginal R^2	Conditional R^2
33	L Cuneus	2.02*	1.72	0.17–3.27	0.060	0.679
148	L ITG	2.09*	2.02	0.24–3.80	0.062	0.679
220	L IFG	2.09*	1.99	0.24–3.76	0.062	0.679
281	R Cuneus	2.12*	1.61	0.24–2.98	0.066	0.679
456	R STS	2.12*	2.0	0.22–3.80	0.061	0.679

Note. For each of the five parcels where neural score predicted behavioral scores with observed beta values at least 1.95 standard deviations above the permuted null distribution, parcel labels, permutation test Z-scores, regression coefficients and their confidence intervals, model intercepts, and marginal and conditional R^2 values are reported.

Figure 6

Quiz scores predicting neural scores (Delta Known-Unknown ASL)



Note. Linear mixed regression results for each of the 117 parcels in the RSA mask are shown on the semi-inflated FSAverage6[62] surface. The scale reflects the Z scores of the observed beta coefficients of each model in the permuted null distribution, displayed as a histogram above the color bar. Of the 117 parcels tested, five met the critical Z threshold for a two-tailed test (1.95), indicated with black outlines. Regression plots and adjusted marginal R^2 values for these five parcels are shown underneath. The overall regression line is plotted as the black line over scatter plots for each recall quiz. The colored lines represent random intercepts for each quiz.

Item Neural Scores Predicting Item Recall

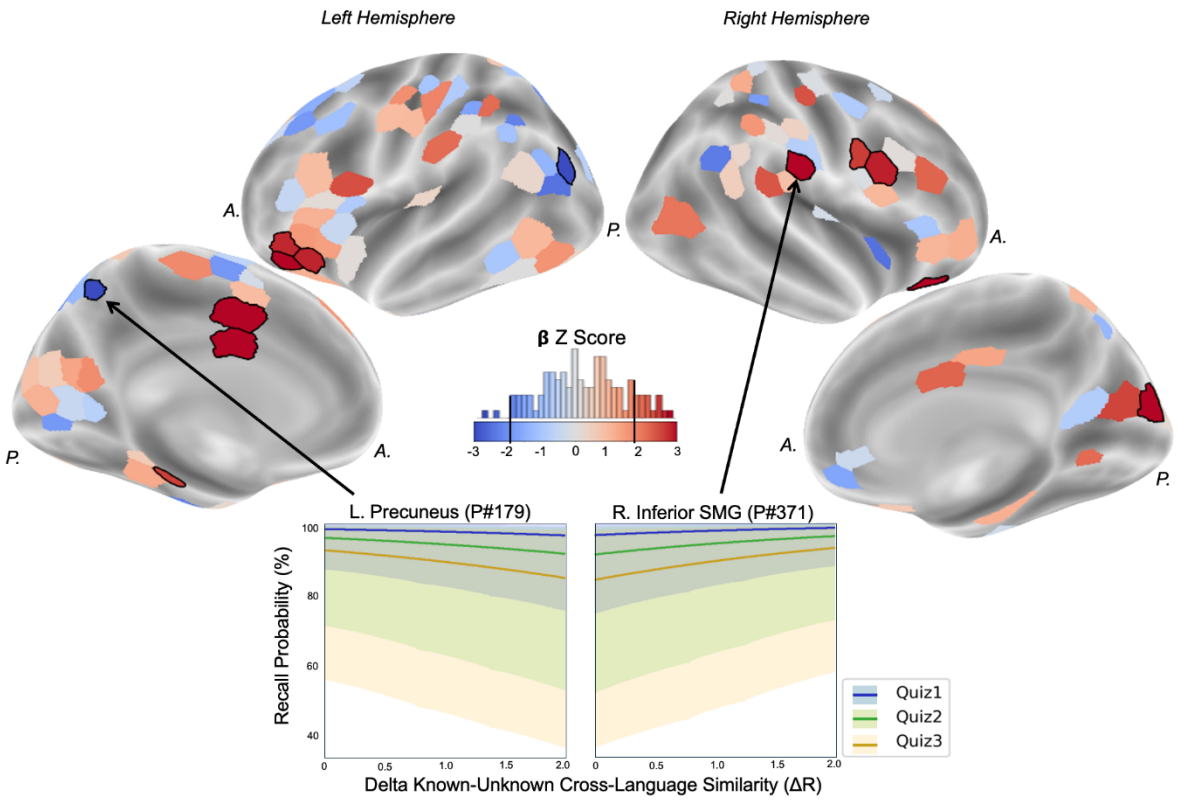
Thus far, we have sought to identify areas where overall correlation between neural scores and semantic relationship between the items predicts overall quiz score, calculated as the proportion of target items that the participant correctly recalled at each time point. However, this design also allows us to compare the neural responses to item recall on the individual stimulus

level. Here, rather than correlation with an a priori model, we compute neural scores directly from the pairwise distances between the items across language (i.e., the representational distance between the same item presented in ASL and in English). If an ASL sign successfully cues an underlying amodal semantic concept, we might expect the neural response to that ASL clip to be more similar to the response to a clip of the same noun in English, than pairs of homologous clips for which the participant does not know the sign's meaning. Thus, we computed the delta between the correlation distances of each known ASL-English pair and the average of all unknown cross-language pairs for each participant in each of the 117 selected parcels. Because scores on the item level are binary (correctly recalled or not), we assess the relationship between those recall scores and cross-language neural correlation for each item by fitting a logistic mixed-effects regression model at each parcel. The resulting coefficient can be interpreted as an odds ratio for recall of a given item. Again, time points 1-3 and participant identity were modeled as random effects, while "item neural score" was modeled as a fixed effect predicting item recall. These models were submitted to the same 1,000-iteration permutation testing procedure as the previous.

Thirteen parcels met the critical Z threshold of 1.95 standard deviations above or below the null distribution, indicating significant prediction of item recall. In eleven of these parcels, representational similarity positively predicted recall – in other words, participants whose neural responses to the same noun in ASL and English were more similar were more likely to recall the correct translation for the sign video. These locations include the left paracentral and pericallosal sulci, left orbital gyrus, left parahippocampal gyrus, right inferior supramarginal gyrus, right inferior gyrus, and right cuneus (parcel 281, which also positively predicted summary quiz scores), highlighted in the heatmap of beta values displayed in Figure 7. Additionally, two parcels (located in the left precuneus, parcel 179, and left middle occipital gyrus, parcel 202) showed a significant negative relationship between ASL-English neural similarity and recall, indicating that greater divergence between the languages predicted successful recall. The full list of locations and model results for each of these parcels are reported in Table 2.

Figure 7

Item recall scores predicting item-level neural scores



Note. Mixed-effects logistic regression results for each parcel in the RSA mask are shown on the semi-inflated FSAverage6[62] surface. Beta estimates for the model estimating item neural score as a predictor of item recall (which included quiz time point, item, and participant as random effects) are shown. Regression results for all 117 tested parcels are shown. Thirteen parcels met the critical Z threshold for a two-tailed test (1.95). For the most predictive positive parcel, located in the right inferior supramarginal gyrus, and the most predictive negative parcel (located in the left precuneus), logistic regression plots are shown below with the modeled probability for each timepoint of the item being recalled given that item's neural score. The slopes of model predictions for each quiz time point are shown as the colored lines, while the shaded area indicates the confidence intervals.

Table 2. Regression results for parcels in which item neural scores best tracked item recall

Parcel	Location	$Z\beta$	Odds Ratio	CI	Marginal R^2	Conditional R^2
130	L Cingulate Sulcus	2.49	1.47	1.08-1.99	0.008	0.554
179	L Precuneus	-2.77	0.67	0.50-0.90	0.009	0.560
182	L Pericallosal Sulcus	2.62	1.49	1.10-2.00	0.008	0.558
202	L Middle Occipital Gyrus	-2.43	0.69	0.51-0.94	0.007	0.555
208	L Orbital Gyrus	2.47	1.44	1.07-1.93	0.007	0.547
209	L Orbital Gyrus	2.17	1.39	1.02-1.88	0.006	0.550
210	L Orbital Gyrus	2.15	1.38	1.02-1.88	0.006	0.550
249	L Parahippo-campal Gyrus	1.98	1.35	1.00-1.82	0.005	0.548
281	R Cuneus	2.29	1.41	1.05-1.90	0.007	0.542
363	R Inf. Precentral Gyrus	2.18	1.39	1.02-1.89	0.006	0.551
364	R. Inf. Precentral Gyrus	2.01	1.36	1.01-1.85	0.005	0.551
371	R Inf. SMG	2.94	1.59	1.16-2.19	0.011	0.549
400	R H-shaped sulcus	2.45	1.42	1.05-1.91	0.007	0.548

Note. For each of the thirteen parcels where item neural score predicted likelihood that a given item would be recalled with observed beta values at least 1.95 standard deviations above or below the permuted null distribution, parcel labels, permutation test Z-scores, regression coefficients and their confidence intervals, marginal R^2, and conditional R^2, are reported.

Discussion

These findings demonstrate that neural patterns reflecting semantic representations of ASL that are both language-dependent and shared with English can be decoded in novice learners after only an hour or two of ASL learning. Moreover, neural scores derived from this decoding analysis reflect understanding on an individual level, and can therefore be an informative measure of successful learning. Consistent with the previous analysis of Study 1, which focused on the decodability of the broad categories (animals, fruits, vehicles) across languages[44], we both identified areas which represented semantic distinctions between the items uniquely for one of the two languages, and at least four areas where brain activity correlated with the semantic model in both languages. These regions included the left superior and middle frontal gyri, which have been associated with semantic judgement tasks[66], including in signed languages[28,67]. The left MFG in particular has been found to increase in univariate activation in non-signers after brief sign language training[68], although MFG activation in novice spoken language learners may be inversely correlated with proficiency as the need for frontal control of semantic processing is reduced upon further development of fluency[69]. Other areas of overlap included the right superior precentral sulcus, and the right superior temporal sulcus (STS).

Beyond simply identifying areas of commonality in semantic decoding, we also demonstrate that the *degree* of neural pattern similarity to the semantic model for studied compared to unstudied vocabulary correlates with behavioral recall scores in areas associated with language processing. The left IFG, left ITG, bilateral cuneus and right STS exhibited neural scores that were predictive of behavioral scores across the three quizzes. In each of these parcels, participants whose neural activity was more strongly correlated with the semantic model for the ASL signs they had studied reliably performed better on the three recall quizzes. This counterbalanced design allows us to use each participant's own unknown trial data as a baseline, ensuring that any correlation to the model measured in the known trials is the direct result of learning. No parcels exhibited a negative relationship between correspondence with the semantic model and quiz score (i.e., where the model performance exceeded chance, increasing neural correlation to the semantic model always positively predicted quiz scores). Of the five significant parcels, one (parcel 456; right STS) was also identified as an area of overlap between the English and Known ASL group-level RSAs, and the other four were identified by the English RSA only. The contribution of English-specific parcels to individual differences that track learning highlights the role of existing semantic representations from a well-known language in supporting new vocabulary learning.

We also assessed whether the overlap of ASL and English (i.e. the representational distance between a known ASL sign and its English homologue, compared to the average of unknown ASL-English pairs for each participant) predicted the likelihood of that participant correctly recalling that individual item. Unlike the summary neural scores, this procedure does not compare neural activity to any *a priori* model. Instead, it directly contrasts the neural

responses to the same noun cued by each language. Strikingly, this cross-language distance positively predicted item recall in a number of areas including the left paracentral sulcus, left orbital gyrus, left parahippocampal gyrus, right inferior supramarginal gyrus, and right cuneus. In two parcels, located in the left precuneus and middle occipital gyrus, ASL-English distance was negatively associated with recall, indicating that participants for whom neural responses were more divergent between languages in these areas were more likely to recall the item. Interestingly, this middle occipital gyrus parcel (parcel 202) was significantly correlated with the phonological model for the unknown signs (see Supplementary Figure 1). It is therefore possible that these areas are representing visual features of the signs which are relied on less by the participants who can readily access the English translations for the video.

Many of the identified regions have been associated with linguistic functions. In particular, the left IFG has long been implicated in language processing, especially comprehension[70]. Consistent with studies of spoken language, it has been shown to support processing of both sign and speech in bimodal bilinguals[23–25,27,28]. Furthermore, univariate activation in this area is greater in signers compared to nonsigners viewing signed stimuli[34], increases within subjects after sign language training[38,71], and has been demonstrated to correlate with proficiency in novice sign learners[36,72]. Some work has also suggested that the left IFG supports learning of spatial-visual properties of new orthographic characters (as shown in English monolinguals after instruction in Chinese characters), although activity in this region may decrease with increasing proficiency over several weeks[73]. These results might seem somewhat contradictory: in some cases, recruitment of left frontal control regions during comprehension has been associated with better performance, and other times with worse. It seems that later in the learning trajectory[73,74], reliance on frontal control can be a hallmark of poorer automatization and greater effort during a language task, but in the earliest stages of learning (e.g. Johnson and colleagues[36], as well as the present work), it can indicate appropriately effortful engagement with the task. This is convergent with work by MacSweeney and colleagues (2004) which found that in deaf native signers, viewing linguistically structured but meaningless pseudosigns activated left IFG more than familiar BSL signs, possibly reflecting effortful lexical search. It is possible that with further exposure, novice learners may show a decrease in left IFG activity, and respond more similarly to the native signers viewing real signs studied by MacSweeney and colleagues[75], but in this brief training paradigm of 1-3 short lessons, it is unsurprising that semantic representation in the left IFG appears to correlate with ability to correctly retrieve the English translations for the studied sign videos.

Behavioral scores also predicted neural scores in the left ITG, an area which is thought to represent amodal concept knowledge and play a role in distinguishing between objects[1]. In a monolingual context, this area has been previously implicated in coding for general (rather than feature-based) semantic similarity between written words[76]. Our findings suggest that this role is not sensitive to the modality of language input. Previous work in novice sign learners has also implicated the ITG in successful sign language learning[71]. Other areas of significant prediction

included the cuneus – where activity predicted both summary quiz scores and item level recall, and which has been shown to increase in univariate activation with spoken language training[69]– and the right STS. The same right STS parcel was correlated with the semantic model in both ASL and English. Prior research suggests this area is involved in auditory processing of speech[77] and biological motion processing[78]. For signed languages in particular, right STS may support processing of prosody and facial expression[79]. For each of these areas which support linguistic functions, our findings demonstrate that not only do multivariate patterns of responses to newly-learned ASL nouns contain information about the semantic relationships between those nouns, but the decodability of this information is predictive of comprehension measured behaviorally.

One final note of interest arising from these neural score analyses was that we observed several areas where item-level language similarity (the representational distance between homologous items in ASL and English, regardless of any a priori semantic model) was meaningfully related to the probability that a participant would recall that specific item. On the contrary, Evans and colleagues[20] concluded that in bilinguals, semantic representations for sign and speech are partially shared only at the level of coarse semantic categories. They found that although individual item identity could be decoded within each language, there was no evidence of fine-grained item representations shared across languages. Our results indicate that in several areas including the right supramarginal gyrus, the representational similarity of the ASL sign and corresponding English word positively predicted recall (two areas, the left precuneus and middle occipital gyrus, also negatively predicted recall, meaning that more divergent neural responses to ASL and English here predicted better performance). One possible explanation for the greater decodability of item-level information in novices compared to bilinguals is that less proficient learners may rely more on existing representations of their native language to scaffold processing of subsequently acquired languages. Evidence from late learners of spoken languages[17,37] suggests that later and less proficient bilinguals exhibit greater neural overlap between languages. Indeed, after the short training paradigm of the present studies, it is unlikely that participants had developed fully nuanced and diverging representations of ASL signs. Studies of novice spoken language learners have shown that the extent of neural similarity between their newly-learned language and native language increases with training and may correlate with proficiency[18,80,81]. Our findings go one step further, demonstrating that not just univariate overlap but subtle shifts in multivariate patterns correlate with individual ability to explicitly translate the semantic content of each sign to a familiar language.

The results of these two studies show that frontal, temporal and occipital areas, especially the right STS, shift to represent semantic information about newly-learned sign language stimuli after very brief learning. The decodability of these shifts can differentiate the novices who learned the most from those who did not know the meanings of the signs both within-group (neural score correlation with individual-level recall performance) and compared to a control group which had learned a different set of stimuli. By assessing the correlation of representational distance between ASL and English with individual recall, we demonstrate that at

this early stage of acquisition, semantic retrieval involves a high degree of overlap between the two languages especially in frontal control regions such as the left IFG. This was true across two different data-driven measures of semantic representation across languages, one which was derived from the correlation of neural activity patterns to an a priori semantic model, and one which directly probed the representational distance between a known word and its homologue in the newly-learned language. This is consistent with other studies of novice signers[71,82] and fluent signers viewing meaningless signs[75], but suggests that novices may rely on their established language more than bilinguals or advanced learners for whom prior work has found that cross-language decoding may only be possible at the broad category level[20]. Instead, novice participants show clear overlap of semantic processing across languages at the individual item level.

While our participants were very novice learners (1-3 hours of training on concrete nouns with no instruction in grammar or other high-level information), future research may explore neural similarity between languages over many more points along the learning trajectory, to assess when or whether representations of the newly-learned language begin to diverge from the established familiar one and become more “expert-like”. Using these and related methods, future research can probe overlapping and diverging representations of language in learners at varying stages of language acquisition to yield deeper insight into how the acquisition of signed language evolves in the brains of adult learners.

References

1. Fairhall, S. L. & Caramazza, A. Brain Regions That Represent Amodal Conceptual Knowledge. *J. Neurosci.* **33**, 10552–10558 (2013).
2. Shinkareva, S. V., Malave, V. L., Mason, R. A., Mitchell, T. M. & Just, M. A. Commonality of neural representations of words and pictures. *NeuroImage* **54**, 2418–2425 (2011).
3. Comstock, L. & Oliver, B. A meta-analysis of task-based differences in bilingual L1 and L2 language networks. Preprint at https://doi.org/10.1101/2021.12.28.474335 (2021).
4. Liu, H. & Cao, F. L1 and L2 processing in the bilingual brain: A meta-analysis of neuroimaging studies. *Brain and Language* **159**, 60–73 (2016).
5. Honey, C. J., Thompson, C. R., Lerner, Y. & Hasson, U. Not Lost in Translation: Neural Responses Shared Across Languages. *J Neurosci* **32**, 15277–15283 (2012).
6. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience* **2**, (2008).
7. Devereux, B. J., Clarke, A., Marouchos, A. & Tyler, L. K. Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects. *J. Neurosci.* **33**, 18906–18916 (2013).
8. Mitchell, T. M. *et al.* Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science* **320**, 1191–1195 (2008).
9. Deniz, F., Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *J. Neurosci.* **39**, 7722–7736 (2019).
10. Liuzzi, A. G. *et al.* Cross-modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage* **150**, 292–307 (2017).
11. Regev, M., Honey, C. J., Simony, E. & Hasson, U. Selective and Invariant Neural Responses to Spoken and Written Narratives. *J. Neurosci.* **33**, 15978–15988 (2013).

12. Simanova, I., Hagoort, P., Oostenveld, R. & van Gerven, M. A. J. Modality-Independent Decoding of Semantic Information from the Human Brain. *Cerebral Cortex* **24**, 426–434 (2014).
13. Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M. & Just, M. A. Identifying bilingual semantic neural representations across languages. *Brain and Language* **120**, 282–289 (2012).
14. Correia, J. *et al.* Brain-Based Translation: fMRI Decoding of Spoken Words in Bilinguals Reveals Language-Independent Semantic Representations in Anterior Temporal Lobe. *J. Neurosci.* **34**, 332–338 (2014).
15. Van de Putte, E., De Baene, W., Price, C. J. & Duyck, W. "Neural overlap of L1 and L2 semantic representations across visual and auditory modalities: a decoding approach". *Neuropsychologia* **113**, 68–77 (2018).
16. Li, H. *et al.* The depth of semantic processing modulates cross-language pattern similarity in Chinese–English bilinguals. *Human Brain Mapping* **44**, 2085–2098 (2023).
17. Cao, F., Tao, R., Liu, L., Perfetti, C. A. & Booth, J. R. High Proficiency in a Second Language is Characterized by Greater Involvement of the First Language Network: Evidence from Chinese Learners of English. *Journal of Cognitive Neuroscience* **25**, 1649–1663 (2013).
18. Ou, J., Li, W., Yang, Y., Wang, N. & Xu, M. Earlier second language acquisition is associated with greater neural pattern dissimilarity between the first and second languages. *Brain and Language* **203**, 104740 (2020).
19. Leonard, M. K. *et al.* Neural stages of spoken, written, and signed word processing in beginning second language learners. *Frontiers in Human Neuroscience* **7**, (2013).
20. Evans, S., Price, C. J., Diedrichsen, J., Gutierrez-Sigut, E. & MacSweeney, M. Sign and Speech Share Partially Overlapping Conceptual Representations. *Current Biology* **29**, 3739-3747.e5 (2019).
21. Kriegeskorte, N. *et al.* Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron* **60**, 1126–1141 (2008).
22. MacSweeney, M., Capek, C. M., Campbell, R. & Woll, B. *The Signing Brain: The Neurobiology of Sign Language*. *Trends in Cognitive Sciences* vol. 12 (2008).

23. Emmorey, K., Mehta, S. & Grabowski, T. J. The neural correlates of sign versus word production. *NeuroImage* **36**, 202–208 (2007).

24. Emmorey, K., McCullough, S., Mehta, S. & Grabowski, T. J. How sensory-motor systems impact the neural organization for language: direct contrasts between spoken and signed language. *Front. Psychol.* **5**, (2014).

25. Horwitz, B. *et al.* Activation of Broca's area during the production of spoken and signed language: a combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia* **41**, 1868–1876 (2003).

26. Hu, Z. *et al.* Brain activations associated with sign production using word and picture inputs in deaf signers. *Brain and Language* **116**, 64–70 (2011).

27. Kassubek, J., Hickok, G. & Erhard, P. Involvement of classical anterior and posterior language areas in sign language production, as investigated by 4 T functional magnetic resonance imaging. *Neuroscience Letters* **364**, 168–172 (2004).

28. Sakai, K. L., Tatsuno, Y., Suzuki, K., Kimura, H. & Ichida, Y. Sign and speech: Amodal commonality in left hemisphere dominance for comprehension of sentences. *Brain* **128**, 1407–1417 (2005).

29. Blanco-Elorrieta, E., Kastner, I., Emmorey, K. & Pylkkänen, L. Shared neural correlates for building phrases in signed and spoken language. *Scientific Reports* **8**, (2018).

30. Braun, A. R., Guillemin, A., Hosey, L. & Varga, M. The neural organization of discourse: An H215O-PET study of narrative production in English and American sign language. *Brain* **124**, 2028–2044 (2001).

31. Emmorey, K., Xu, J. & Braun, A. Neural responses to meaningless pseudosigns: Evidence for sign-based phonetic processing in superior temporal cortex. *Brain and Language* **117**, 34–38 (2011).

32. MacSweeney, M. *et al.* Lexical and sentential processing in British Sign Language. *Human Brain Mapping* **27**, 63–76 (2006).

33. Banaszkiewicz, A., Costello, B. & Marchewka, A. Early language experience and modality affect parietal cortex activation in different hemispheres: Insights from hearing bimodal bilinguals. *Neuropsychologia* **204**, 108973 (2024).

34. Li, Q., Xia, S., Zhao, F. & Qi, J. Functional changes in people with different hearing status and experiences of using Chinese sign language: An fMRI study. *Journal of Communication Disorders* **50**, 51–60 (2014).

35. Newman, A. J., Bavelier, D., Corina, D., Jezzard, P. & Neville, H. J. A critical period for right hemisphere recruitment in American Sign Language processing. *Nat Neurosci* **5**, 76–80 (2002).

36. Johnson, L. *et al.* Functional Neuroanatomy of Second Language Sentence Comprehension: An fMRI Study of Late Learners of American Sign Language. *Front. Psychol.* **9**, (2018).

37. Tan, L. H. *et al.* Neural systems of second language reading are shaped by native language. *Human Brain Mapping* **18**, 158–166 (2003).

38. Williams, J. T., Darcy, I. & Newman, S. D. Bimodal bilingualism as multisensory training?: Evidence for improved audiovisual speech perception after sign language exposure. *Brain Research* **1633**, 101–110 (2016).

39. Chao, L. L., Weisberg, J. & Martin, A. Experience-dependent Modulation of Category-related Cortical Activity. *Cerebral Cortex* **12**, 545–551 (2002).

40. Weisberg, J., van Turennout, M. & Martin, A. A Neural System for Learning about Object Function. *Cerebral Cortex* **17**, 513–521 (2007).

41. Cetron, J. S. *et al.* Decoding individual differences in STEM learning from functional MRI data. *Nature Communications* **10**, 2027 (2019).

42. Cetron, J. S. *et al.* Using the force: STEM knowledge and experience construct shared neural representations of engineering concepts. *npj Sci. Learn.* **5**, 6 (2020).

43. Mason, R. A. & Just, M. A. Neural Representations of Physics Concepts. *Psychological Science* **27**, 904–913 (2016).

44. Hillis, M. E. *et al.* Overlapping semantic representations of sign and speech in novice sign language learners. *Proceedings of the Annual Meeting of the Cognitive Science Society* **44**, (2022).

45. Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M. & Emmorey, K. ASL-LEX: A lexical database of American Sign Language. *Behav Res* **49**, 784–801 (2017).

46. Sehyr, Z. S., Caselli, N., Cohen-Goldberg, A. M. & Emmorey, K. The ASL-LEX 2.0 Project: A Database of Lexical and Phonological Properties for 2,723 Signs in American Sign Language. *The Journal of Deaf Studies and Deaf Education* **26**, 263–277 (2021).

47. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient Estimation of Word Representations in Vector Space. in (2013).

48. Yamada, I. *et al.* Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia. Preprint at https://doi.org/10.48550/arXiv.1812.06280 (2020).

49. Brentari, D. *A Prosodic Model of Sign Language Phonology*. (MIT Press, 1998).

50. Campbell, R., Martin, P. & White, T. Forced Choice Recognition of Sign in Novice Learners of British Sign Language. *Applied Linguistics* **13**, 185–201 (1992).

51. Lieberth, A. K. & Gamble, M. E. B. The role of iconicity in sign language learning by hearing adults. *Journal of Communication Disorders* **24**, 89–99 (1991).

52. Akers, E. M., Midgley, K. J., Holcomb, P. J., Meade, G. & Emmorey, K. Neural effects differ for learning highly iconic versus non-iconic signs in hearing adults. *Bilingualism: Language and Cognition* 1–13 (2023) doi:10.1017/S1366728923000809.

53. Baus, C., Carreiras, M. & Emmorey, K. When does iconicity in sign language matter? *Language and Cognitive Processes* **28**, 261–271 (2013).

54. Ortega, G. & Morgan, G. The effect of iconicity in the mental lexicon of hearing non-signers and proficient signers: evidence of cross-modal priming. *Language, Cognition and Neuroscience* **30**, 574–585 (2015).

55. Bosworth, R. G. & Emmorey, K. Effects of iconicity and semantic relatedness on lexical access in american sign language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **36**, 1573–1581 (2010).

56. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res* **51**, 195–203 (2019).

57. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111–116 (2019).

58. Esteban, O. *et al.* Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nat Protoc* **15**, 2186–2202 (2020).

59. Gorgolewski, K. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front. Neuroinform.* **5**, (2011).

60. RRID | Resource Report (RRID: SCR_001362). https://rrid.site/data/record/nlx_144509-1/SCR_001362/resolver?q=nilearn&l=nilearn&i=rrid:scr_001362.

61. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095–3114 (2018).

62. Fischl, B. FreeSurfer. *NeuroImage* **62**, 774–781 (2012).

63. Jolly, E. Pymer4: Connecting R and Python for Linear Mixed Modeling. *JOSS* **3**, 862 (2018).

64. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. Preprint at https://doi.org/10.48550/arXiv.1406.5823 (2014).

65. Giacalone, M. & Alibrandi, A. Overview and Main Advances in Permutation Tests for Linear Regression Models. *JMSS* **5**, (2015).

66. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex* **19**, 2767–2796 (2009).

67. San José-Robertson, L., Corina, D. P., Ackerman, D., Guillemin, A. & Braun, A. R. Neural systems for sign language production: Mechanisms supporting lexical selection, phonological encoding, and articulation. *Human Brain Mapping* **23**, 156–167 (2004).

68. Alotaibi, S., Alamri, S., Alsaleh, A. & Meyer, G. Neural adaptations in short-term learning of sign language revealed by fMRI and DTI. *Sci Rep* **15**, 5345 (2025).

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

70. Geschwind, N. Specializations of the human brain. *Scientific American* **241**, 180–199 (1979).

71. Banaszkiewicz, A. *et al.* Multimodal imaging of brain reorganization in hearing late learners of sign language. *Human Brain Mapping* **42**, 384–397 (2021).

72. Newman-Norlund, R. D., Frey, S. H., Petitto, L. A. & Grafton, S. T. Anatomical substrates of visual and auditory miniature second-language learning. *Journal of Cognitive Neuroscience* **18**, 1984–1997 (2006).

73. Deng, Y., Booth, J. R., Chou, T.-L., Ding, G.-S. & Peng, D.-L. Item-specific and generalization effects on brain activation when learning Chinese characters. *Neuropsychologia* **46**, 1864–1876 (2008).

74. Williams, J. T., Darcy, I. & Newman, S. D. Neural substrates of sign language vocabulary processing in less-skilled hearing M2L2 signers: Evidence for difficult phonological movement perception. *Bilingualism* **21**, 550–562 (2018).

75. MacSweeney, M. *et al.* Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage* **22**, 1605–1618 (2004).

76. Liuzzi, A. G., Aglinskas, A. & Fairhall, S. L. General and feature-based semantic representations in the semantic network. *Sci Rep* **10**, 8931 (2020).

77. Kriegstein, K. V. & Giraud, A.-L. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage* **22**, 948–955 (2004).

78. Sokolov, A. A. *et al.* Biological motion processing: The left cerebellum communicates with the right superior temporal sulcus. *NeuroImage* **59**, 2824–2830 (2012).

79. Newman, A. J., Supalla, T., Hauser, P. C., Newport, E. L. & Bavelier, D. Prosodic and narrative processing in American Sign Language: An fMRI study. *NeuroImage* **52**, 669–676 (2010).

80. Leonard, M. K. *et al.* Language Proficiency Modulates the Recruitment of Non-Classical Language Areas in Bilinguals. *PLOS ONE* **6**, e18240 (2011).

81. Stein, M. *et al.* Reduced frontal activation with increasing 2nd language proficiency. *Neuropsychologia* **47**, 2712–2720 (2009).

82. Williams, J. T., Darcy, I. & Newman, S. D. Modality-specific processing precedes amodal linguistic processing during L2 sign language acquisition: A longitudinal study. *Cortex* **75**, 56–67 (2016).