Insights into Early Word Comprehension - Tracking the Neural Representations of Word Semantics in Infants

by

Rohan Saha

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science
University of Alberta

Abstract

Infants start developing rudimentary language skills and can start understanding simple words well before their first birthday [1]. This development has also been shown primarily using Event Related Potential (ERP) techniques to find evidence of word comprehension in the infant brain [2, 3]. While these works validate the presence of semantic representations of words (word meaning) in infants, they do not tell us about the mental processes involved in the manifestation of these semantic representations or the content of the representations. To this end, we use a *decoding* approach where we employ machine learning techniques on Electroencephalography (EEG) data to predict the semantic representations of words found in the brain activity of infants. We perform multiple analyses to explore word semantic representations in two groups of infants (9-month-old and 12-month-old). Our analyses show significantly above chance decodability of overall word semantics, word animacy, and word phonetics. As we analyze brain activity, we observe that participants in both age groups show signs of word comprehension immediately after word onset, marked by our model's significantly above chance word prediction accuracy. We also observed strong neural representations of word phonetics in the brain data for both age groups, some likely correlated to word decoding accuracy and others not. Lastly, we discover that the neural representations of word semantics are similar in both infant age groups. Our results on word semantics, phonetics, and animacy decodability, give us insights into the evolution of neural representation of word meaning in infants.

Keywords: Machine Learning, Brain Decoding, EEG, Semantics

Preface

A major part of this thesis has been published as a preprint on bioRxiv.org [4]. We will submit the results as an extended paper to a journal.

To my parents Rajendra and Manika, and my sister Rashmika, for always being there for me.

Acknowledgements

I would like to thank **my supervisor, Dr. Alona Fyshe** for her constant support, guidance, and impeccable mentorship. She made sure I was exposed to new ideas and made me a better researcher.

I am also thankful to **Dr. Janet Werker** and **Dr. Jennifer Campbell** for their valuable inputs during group meetings in all stages of the project.

My **lab-mates** were extremely kind in providing suggestions and insights during discussions.

Finally, I am truly grateful to **my family** for their continuous love and support throughout my academic career.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Human beings have a marvellous capability of using language to communicate and understand each other. This extraordinary skill starts developing early on in one's life and continues to evolve as we grow up. When we hear or read words, our brain processes this information and provides us with a semantic understanding of these words. But how does our brain process this information, and how does it map the text that we read or a word that we hear to its meaning? That question is of paramount interest as it gives us a preview of the inner workings of the human mind. Accordingly, many studies have explored the neural responses (brain-activity) to word stimuli [5–9]. These neural responses to stimuli have primarily been explored in adults but less studied in infants. The present work examines the neural responses recorded from infants from two age groups while they were presented with a set of word stimuli.

Each word in a language has some *meaning* associated with it. This meaning is usually referred to as *semantics* and represents the conceptual ideas attached to each word (e.g., cat is an animal, apple is a fruit and is edible). These semantics can be modeled computationally using mathematical models obtained from statistical modeling of large corpora of text. The trained semantic models of text express various characteristics of words such as gender, plurality, etc. [10]. These semantic models have been beneficial for studying language processing in the brain.

One can then put together recorded brain activity and semantic models to study the processes involved in the brain during word understanding. We can then use a *decoding* approach where a machine learning model is used to predict the stimuli from recorded brain activity.

Previous studies conducted on brain-imaging data recorded from adults have shown success in decoding word semantics [5, 11, 12]; we extend this idea to brain-imaging data recorded from infants. Specifically, we analyze the change in neural response patterns as the infants hear single words. Infants in two age groups (9-month-old and 12-month-old) listened to single words spoken by a native English speaker while their EEG data was recorded. We then trained a machine learning model to decode the semantic representations of the words from the brain activity.

Our analyses on brain-imaging data collected from infants provide evidence that,

- EEG data recorded from 9-month-old and 12-month-old infants can be used with a machine learning approach to reveal semantic information.

- Both 9-month-old and 12-month-old infants show high decoding accuracy of words immediately after word onset.

- Individual word stimuli properties such as phonetics can be decoded with high accuracy from both 9-month-old and 12-month-old infants.

- The neural responses of 9-month-old and 12-month-old infants show similar representations of word semantics.

In the following chapters, we will look at existing literature in the domain of semantic research, the methodology used for our analyses, and the results.

1.1 Thesis Organization

In this thesis, we apply computational techniques to EEG data to explore infants' neural responses as they process word stimuli. The next chapter provides background work that will help situate the work presented here. We first cover studies that use machine learning techniques for brain decoding and describe computational models for representing word semantics. We also discuss the existing semantic research in infants. Chapter 3 outlines the design of our study and the various components involved. We describe the stimuli, the data collection process, the prediction framework, performance metrics, and the statistical significance testing methodology. In chapter 4, we report our findings and discuss the results obtained. Finally, in chapter 5, we conclude by recapitulating our study, summarizing our results, and considering future directions.

Chapter 2

Background

In this chapter, we first provide a short overview of machine learning techniques for brain-imaging related analyses, then explain computational models for word semantics, and finally discuss existing research in representations of single word semantics in the human brain. Throughout the rest of this thesis, we use computational models of word semantics and machine learning to study word meaning comprehension in the infant brain.

2.1 Machine Learning for brain decoding

Machine Learning has been gaining ground as one of the primary analysis techniques for neuroscience research. With the advent of convenient and easy-to-use software packages, there is no hindrance in considering machine learning for neuroimaging analysis [13]. A machine learning model is usually a mathematical function $h(X)$ that maps a set of inputs X to a set of outputs Y. Effectively, h learns a transformation which, when applied to the input data X, results in predictions Y. The function h is generally learned by optimizing an objective function (loss function) using an optimization algorithm. Under the umbrella of machine learning, multiple algorithms exist, each suitable for a different task. They can be applied to any type of brain-imaging data (fMRI, MEG, EEG, fNIRS etc.) [14, 15].

Functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and Electroencephalography (EEG) are imaging modalities that record brain-activity. fMRI is a non-invasive brain-imaging technique that records brain-activity by measuring changes in blood oxygenation level as the participant is engaged in information processing. It provides excellent spatial resolution and is ideal for voxel-level analyses. fMRI also has the advantage of recording the whole brain data as a three-dimensional image, thus providing insights into various regions of the brain. However, fMRI data is expensive to collect and suffers from poor temporal resolution. fNIRS also measures changes in blood oxygenation level but has higher temporal resolution than fMRI. On the other hand, techniques such as Magnetoencephalography (MEG) and Electroencephalography (EEG) are more suited for temporal analyses of neural responses. In our study, we use EEG data as it has the advantage of high temporal resolution. EEG is a non-invasive technique because it records brain activity in the form of electrical signals from electrodes placed on the participants' scalp. EEG also requires minimal experimental set-up, and is less expensive to collect. Therefore, it is an ideal brain-imaging technique when participants tend to move a lot (e.g., infants).

Machine Learning can be applied to brain-imaging data in many ways. Among these, the *encoding* and *decoding* approaches have been successfully used to study language in the human brain. A decoding approach is one where the stimulus is predicted from the neural responses [15, 16] and an encoding approach is one where the neural responses are predicted from the stimuli [8]. A decoding problem can be used to see if the brain-imaging data contains information for a stimuli set. On the other hand, an encoding approach can be helpful to understand the computational rules responsible for the neural response given a stimulus. In this work, we use the decoding approach where we train a machine learning model on brain-imaging data collected from infants to predict the stimuli presented to them.

2.2 Computational Models of Semantic Representation of Words

Every word has a meaning attached to it which allows us to comprehend the overall meaning of, for example, a sentence. While humans have a mental representation of word meaning, computers do not know the meaning attached to the words. Fortunately, each word can be represented in a numerical format, allowing computers to have a representation of word meaning. To put it another way, attaching semantic meaning to words can be accomplished by establishing a numerical representation for each. This can be achieved by assigning each word in a vocabulary a *vector* of real numbers called a *word vector*. Each word vector is a list of multiple real values, which effectively represents the word meaning in a higher dimensional numerical space. These word vectors are typically built using the words' co-occurrence statistics with other words. This leverages the concept of distributional semantics, which means that words with similar meanings frequently occur in similar contexts [17] (for example, the word 'orange' frequently occurs with 'fruit' than with 'mountain'). Therefore, similar words have similar real-valued *word vectors* and thus are close to each other in the higher dimensional space. Due to this property of word vectors, they are useful for representing word semantics in computers, i.e., word meaning. Specifically, semantics for a word can be represented by its equivalent word vector.

Word2Vec is a fantastic example of a modern word vector model introduced in [18], that provided two neural network based algorithms for learning vector representations of words. These were the Continuous Bag of Words (CBOW) and Skip-gram models, both of which were shown to provide high performance on Natural Language Processing (NLP) based tasks [19]. In general, word vector models provide n-dimensional vectors of numbers for each word. These word vectors can then help represent the semantics of the stimuli words in downstream tasks such as decoding. In the work laid out in this thesis, we will use the word vectors from the Skip-gram model of

Word2Vec to represent the word semantics in a numerical format and then decode these word vectors from the brain-imaging data.

2.3 Semantic Research using brain-imaging data

2.3.1 Semantic Decoding in Adults

Semantic representation of words in the human brain has been studied fairly extensively and new discoveries are constantly being made. Current brain-imaging techniques provide us with a lot of data for finding patterns that lead to ideas and insights. However, such studies have usually been conducted on brain imaging data collected from adults. This is primarily because collecting data from adults is generally seamless and easy.

To decode these semantic representations from the brain, machine learning presents itself as one of the major analyses techniques. A cognitive state classification conducted by Mitchell *et al.* in 2004 [20] showed that machine learning methods can be employed for decoding semantic categories of written stimuli from whole brain fMRI (functional magnetic resonance imaging) data collected from adults. In this study, university students participated in various tasks while their fMRI data was collected. Then a comparative analysis was performed using different machine learning algorithms (Gaussian Naïve Bayes, Support Vector Machines and k-Nearest Neighbors) on the fMRI data to classify cognitive states. This indicated that machine learning models can be successfully applied to brain-imaging data .

Other studies that required participants to perform a language related task, also showed that the word semantics information is encoded in the brain. A study conducted in 2014 by Fyshe *et al.* built vector space models of words by incorporating brain-imaging data [21]. Eighteen participants took part in the study and their MEG and fMRI data were recorded while they viewed sixty concrete nouns. To represent these sixty nouns, vector space models that encode semantic information were built.

A simple L_2 regularized linear regression model was used to predict these word vectors from brain-imaging data. It was observed that when brain-imaging data and text were leveraged to build the word vectors, the regression model outperformed in predicting the word vectors compared to predicting them when the word vectors were built using only text. The finding that the word vectors built by leveraging brain-imaging data in addition to text were better at representing word semantics than text alone shows that the brain-imaging data had semantics encoded in it.

Another study conducted by Sudre *et al.* in 2012 used MEG data where machine learning was used to decode perceptual and semantic features of 60 concrete nouns [9] using a classification approach. By decoding perceptual and semantic features, an MEG based classifier was able to determine two different concrete nouns not seen during training. The temporal analyses conducted provided insights into the difference between the time course of MEG magnitude and the decodable semantic information; it showed that perceptual information can be decoded before semantic information. This study showed that applying machine learning to timeseries data can help us discover dynamic patterns that vary with time.

Machine Learning has also been used with MEG data to decode phrase stimuli semantics from brain-imaging data. For example, Fyshe *et al.* in 2019 showed that a simple linear model (ridge regression) was able to track the neural decoding of adjective and noun phrase semantics from MEG data [5]. A total of thirty adjective-noun phrases were present in the stimuli set, out of which twenty-eight were used for training and two for model performance evaluation. The decoding model also predicted the semantics of the adjective-noun phrases, indicating that neural responses contain information pertaining to the stimuli much later after stimuli offset. All in all, machine learning methods are a great choice for semantic decoding research and also for studies involving temporal analyses of brain-imaging data.

2.3.2 Decoding infant brain-activity

As we have seen with brain-imaging data recorded from adults, machine learning can also be applied to brain-imaging data collected from infants. In fact, classical and modern machine learning techniques have been used to investigate EEG data in developmental studies. For example, Gibbon *et al.* used Support Vector Machine (SVM) and Convolutional Neural Network (CNN) to show that rhythmic stimuli can be accurately classified from EEG data collected from infants of 8-weeks-old [22]. Another study used SVM on EEG data obtained from 6-month-old infants to accurately classify their risk of developing language related disorders [23]. This supports the idea that machine learning can be used on brain-imaging data collected from infants in downstream tasks, and using computational models can also provide us with valuable insights into the dataset over traditional methods such as Event Related Potential (ERP).

Other techniques such as the Multivariate Pattern Analysis (MVPA) technique have also been used to decode the infant mind. Studies such as [24] used functional near-infrared spectroscopy (fNIRS) brain-imaging data (a technique that measures changes in cortical blood oxygenation) in a task that classifies a set of stimuli into audio or visual categories. In this study, two datasets were used where in the first dataset, infants viewed either a visual stimulus or an audio stimulus, and in the second dataset, the infants viewed a combination of audio-visual stimuli. fNIRS data was collected while the infants were presented with the stimuli. Then, a correlation-based accuracy metric was used to decode unlabelled stimuli from a held-out test infant's data. Above chance accuracy was observed for both datasets providing evidence that fNIRS can be a useful technique to study the human brain. Using fNIRS and fMRI techniques is ideal when spatial information is vital. On the other hand, our study uses EEG for a language related task as it is ideal for observing changes in neural responses over time.

Compared to the plethora of semantic decoding research conducted with neuroimaging data collected from adults, there is a scarcity of research where machine learning techniques have been applied to decode semantics from infant neuroimaging data. Semantic research in infants has traditionally been carried out using Event Related Potentials (ERP), a characteristic of EEG data [25]. A study conducted in 2012 by Parise *et al.* [2] used ERP data collected from 9-month-old infants to show that the participants can detect a mismatch between an object and a word label that preceded the object suggesting that a semantic link is formed between the word stimuli and object stimuli. Infants aged 9-month-old were presented with live audio stimulus from their mother in one experimental condition and from an experimenter in another condition. Then a congruent or an incongruent object relative to the preceding audio stimulus was shown on a screen placed in front of infants. The infants had seemed to activate the object features associated with familiar words in the mother-speech condition and matched to the image that occurred in front of them, thus indicating that the infants *understood* their mother's speech.

A crucial component in language acquisition is the N400 component of an ERP. The N400 signal has been linked to semantic processing of language [26] and is a key component in early language acquisition. For example, 12-month-old infants have been shown to have the presence of the N400 signal indicating that semantic comprehension occurs as early as one year of age. This was shown in a study conducted Friedrich *et al.* in 2012 [3] where the relation between the N400 ERP signal and the development of behavioral language skills was investigated. A picture-word priming paradigm was set up to examine infants' word production and comprehension skills. With the help of this paradigm, it was found that the N400 signal can be detected in infants who have particularly high early production skills but not detected in infants who have normal language development at 12 months of age. Overall, the ERP technique has been successfully used to confirm the semantic processing of words in infants.

2.4 Conclusion

The plethora of brain-imaging data, semantic word vectors, and compute resources available has allowed machine learning to be a widely used analysis technique. Applying such computational models on brain-imaging data has allowed us to understand how the brain represents language semantics. Although an abundance of research involving neural decoding tracking has been conducted on brain-imaging data collected from adults and semantic research in infants using traditional techniques such as ERP, there is a lack of research where machine learning has been used to track the neural decoding of words on infant brain-imaging data. To address this gap, we extend the idea of tracking semantic neural decoding to infants in this thesis. In the next chapter, we explain the methodology used for this study, after which we discuss our results.

Chapter 3

Methods

3.1 Introduction

Our study focuses on tracking the neural responses to word stimuli using EEG data collected from nine-month-old and twelve-month-old infants. In this chapter, we explain each component of our experimental design framework. We describe the stimuli and their types, the acquisition of the EEG data, the processing steps before running the machine learning prediction framework, the evaluation metric (2 vs 2 accuracy), and significance testing methods.

3.2 Stimuli

The stimuli presented to the participants consisted of two parts, an audio stimulus and a visual stimulus. The audio stimuli represented single words, which were spoken into a microphone by a native English female speaker (the experimenter). The visual stimuli represented images shown on a screen placed in front of the participant.

3.2.1 Audio Stimuli

The word stimuli set consisted of sixteen total words equally divided into animate and inanimate words. The stimuli words were selected using the reported average age of acquisition from the MacArthur–Bates Communicative Development Inventory (CDI) [27], and previous experimental studies [1, 28–31].

Table 3.1 shows the list of sixteen words grouped by animacy. It is useful to note that each spoken word takes a different duration, but requires less than one second to speak.

Stimuli Words	
Animate	**Inanimate**
baby	banana
bear	bottle
bird	cookie
bunny	cracker
cat	cup
dog	juice
duck	milk
mom	spoon

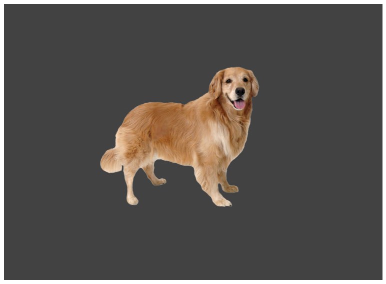
Table 3.1: Stimuli words used in the study.

3.2.2 Visual Stimuli

For each audio stimuli, an image followed. Sixteen images matching the auditory stimuli were presented to the participants. Each image was publicly available on Google Images and had black backgrounds. The images represented the real-life objects equivalent to the corresponding word label (word stimulus) and not toys. For example, the image for the word label *dog* showed a real-life dog and not a toy dog. Figure 3.1 shows an example of an animate and an inanimate image presented to the infants.

3.2.3 Stimuli Presentation

Before each trial, the computer displayed a dynamic video as an attention getter on the screen, which displayed a small rotating asterisk that changed colors. This



(a) Dog as an animate image in the stimuli set.

(b) Cup as an inanimate image in the stimuli set.

Figure 3.1: An example of an animate and inanimate image presented to the participants after the word stimuli.

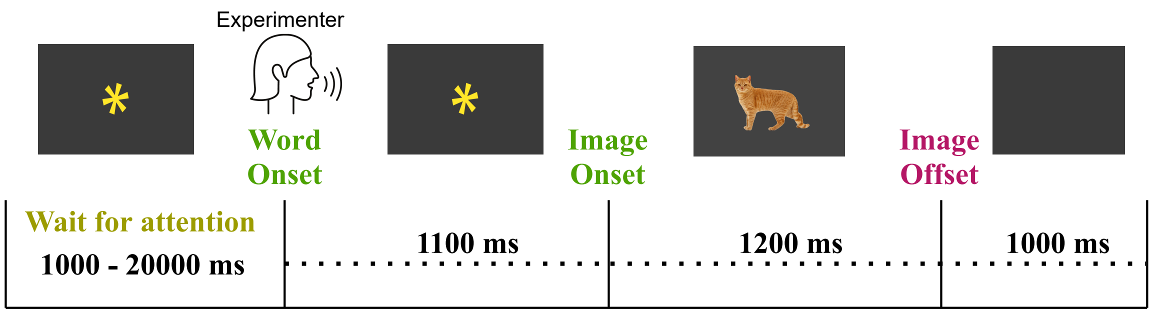


Figure 3.2: Timing diagram showing stimulus presentation design. The word stimulus was in audio format and the image was shown on a screen. Before each trial, a dynamic video was shown with a rotating asterisk that changed colors. After word onset, the rotating asterisk was shown for at least 1100ms. Next, an image appeared on the screen that lasted for 1200ms. After the image offset, a blank screen followed that lasted 1000ms.

video was used to ensure the participants' attention based on the criteria that they were either looking at the rotating asterisk or the experimenter. The experimenter spoke the word marking the word stimuli onset, following which the rotating asterisk was shown for 1100ms. Next, the image was presented on the screen that lasted for 1200ms. Finally, a blank screen was shown for 1000ms (see Figure 3.2 for details). On average, the time between image offset and the next word stimulus varied from 2100ms to 2500ms because participants' attention was required, and the reaction time of the speaker was accounted for to the next word stimulus provided by a computer software.

For each trial, the Eprime software [32] provided a stimulus word through the experimenter's headphones, such that the infants and the parents could not hear the word stimulus for the next trial. Then, as the speaker said the word, the microphone in front of the speaker registered the word stimulus and would signal the Eprime software to present an image after, at least 1100ms after the word onset. For the computer to register the onset of the spoken word stimulus, a threshold level of 65dB was used.

An image was shown for 1200ms on the screen in front of the participant after each word stimulus. For half of the trials, the image matched the word label (for example, an image of *dog* was shown for the preceding word label *dog*). For the other half of the trials, the image did not match the preceding word label (for example, an image of a *cup* was shown for the preceding word label *banana*). Half of the mismatched images matched the preceding word for its animacy category, and the remaining half mismatched the animacy category of the word. This resulted in four cases of word-image pairs. These were:

1. Animates congruent.
2. Animates incongruent.
3. Inanimates congruent.
4. Inanimates incongruent.

In the congruent case, the image matched the animacy group of the word (e.g., animate word *baby* and an image of the animate object *cat*), and in the incongruent case, the animacy did not match the preceding word stimulus (e.g., inanimate word *cup* and an image of the animate object *dog*). The images were presented in a semi-random order such that there could not be more than two presentations of the same word type (animate or inanimate) in a sequence. For our study, because we are trying to analyze the neural responses of word semantics, we will use the data recorded only during the word stimuli presentation for both congruent and incongruent trials.

3.3 Data Acquisition and Processing

A total of 46 infants participated in our study. Of this initial sample, 21 9-month-old (10 female, 11 male) infants were included in the final analysis. Seven additional 9-month-old infants were tested but excluded from the final sample due to fussiness (n=5), poor electroencephalogram (EEG) impedance (n=1), and excessive body movements (n=1). Fourteen 12-month-old (7 female, 7 male) infants were included in the final sample. An additional four 12-month-old infants were tested but excluded due to language criteria (n=1) and fussiness (n=3). All infants were raised in native English speaking households, had no cognitive impairments, and were born full-term.

Infant brain imaging data were collected using Electroencephalography (EEG) at the Infant Studies Centre, University of British Columbia (Ethics ID H19-01411). EEG is a brain-imaging technique that involves placing electrodes on the scalp to record the electrical activity of the brain. It is a non-invasive and affordable method that provides excellent temporal resolution. EEG is ideal for our study as we needed high temporal precision to analyze the changes in brain activity over time.

We used a Geodesic 64 channel EEG cap which had a data capturing frequency of 1000 hertz. Channels 61-64 were removed during pre-processing since these channels corresponded to cheek channels which were not present on all caps. Channel 65 was used as a reference. In addition, we removed bad channels by visually identifying all segments for channels that had comparatively (and consistently) unusually high noise levels or had flat readings (indicating a poor electrode connection). In addition, if there was a cluster of bad channels, we did not include that baby's data in our analysis. We used the EEG data in its raw format for all our experiments with no downsampling as this might lead to data loss. The EEG data were filtered for 0.1 Hz to 50Hz, segmentation for the range -200ms to 1000ms where 0ms marked the onset of the word stimuli, baseline corrected from -200ms to 0ms, and referenced using averaging referencing. We used EGI Netstation 5.4.2 software[1] to record the data, and Eprime was used to present the stimuli and control the timing of the experiment. The experimental procedure was carried out using Eprime software with Chronos hardware [32].

Infants were seated on their parent's lap in front of a computer monitor for the duration of the experiment. Next to the screen was the experimenter who provided the word stimulus (see Figure 3.3). Before initiating a trial, we ensured that participants were relatively still and attentive to either the speaker or the screen before moving on to the next trial.

Figure 3.3: Experimental room setup where the data was collected. The participants sat on their parent's lap and the experimenter provided the word stimuli. Following the word stimuli, an image was shown on the screen placed in front of the participant.

[1]https://www.egi.com/research-division/net-station-eeg-software

In the end, we had a total of 1026 samples of EEG data recorded from 9-month-old infants and a total of 683 samples of EEG data recorded from 12-month-old infants. We also made sure that any misalignment of the EEG data to the word onset was manually corrected.

3.4 Prediction Framework

3.4.1 Word vectors

To obtain a vector representation of word semantics, we used word vectors from the Word2Vec model [18] pretrained on the Google News Dataset (an internal Google dataset containing news articles and comprising about a hundred billion tokens[2]). These word vectors represent the semantic properties of the stimuli. Each word vector has 300-dimensions. For our study, we used the vectors acquired from the skip-gram algorithm, a neural network trained to predict the context for a given word. We chose vectors from Word2Vec because the existing studies show that word vector representations of stimuli can be decoded from brain-imaging data in adults [6, 7, 9, 12, 33, 34].

3.4.2 Prediction Model

Our primary analyses use a simple machine learning algorithm. Specifically, we use the Ridge Regression model from scikit-learn [35], which was trained on the EEG data to predict the word vectors. Ridge regression reduces the weights for the correlated features by driving them to be close to zero, unlike L1 (or Lasso) regression which sets the weights for the correlated features to exactly zero. In addition, ridge regression has a closed-form solution whereas, L1 regression does not.

To train the model, we learn a mapping h from X to Y. h takes the brain-imaging data X and predicts word vectors $\hat{Y}$.

$$h(X) = \hat{Y} \tag{3.1}$$

[2]https://code.google.com/archive/p/word2vec/

or

$$h(X) = X \times \hat{\mathbf{w}} = \hat{Y} \tag{3.2}$$

To learn the mapping h, we estimate the weights $\hat{\mathbf{w}}$ by minimizing the loss function:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||X\mathbf{w} - Y||_2^2 + \lambda \mathbf{w}^T \mathbf{w} \tag{3.3}$$

The term $\lambda \mathbf{w}^T \mathbf{w}$ is called the ridge or the L2 regularizer and the hyperparameter λ controls the regularization strength.

In Figure 3.4, we show the pipeline for the data flow which includes the prediction model.

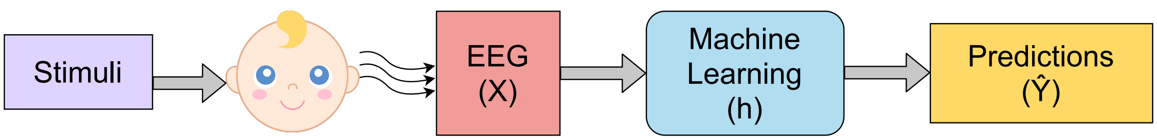


Figure 3.4: Components of the prediction pipeline. EEG data (X) were recorded while the participants heared the the word stimuli. Then a machine learning model (h) was trained on the input EEG data (X) to predict the word vectors (Y) for the word stimuli.

The EEG dataset collected is represented by $X \in \mathbb{R}^{N \times p}$, where N is the number of samples and $p = sensors \times time$, is the number of features. A machine learning model learns a function h that transforms the input EEG data X into a different m dimensional space Y. This v dimensional space consists of the 300-dimensional word vectors that represent the semantics of the word stimuli $Y \in \mathbb{R}^{N \times v}$.

Our data has a total of p = 60 $\times$ 1000 = 60000 features for 60 sensors and 1000 time points for each sensor. We divided our data into equal length chunks of 100ms windows and used a sliding step of 10ms. Therefore, for each window, we trained a ridge regression model, which takes in a data matrix of size $N \times t$, where $t = 60 \times 100 = 6000$ is the total number of features (60000 total data points for 1 second of data chunked into 100ms windows). So our original data matrix $X^{N \times p}$ is now $X^{N \times t}$ ($t < p$) on which the model is trained to predict the 300-dimensional word vectors (see Figure 3.5).

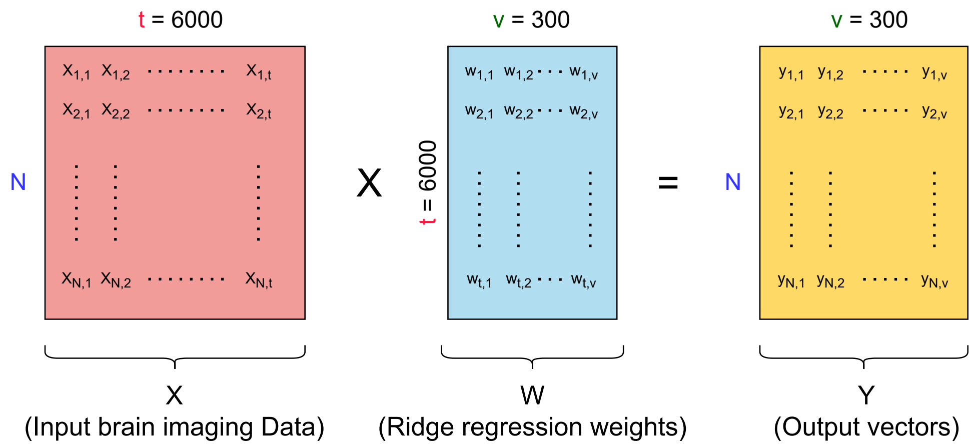


Figure 3.5: The Ridge Regression predicts the 300-dimensional word vectors ($N \times v$) by multiplying the input brain-imaging data matrix X (size $N \times t$) with the weight matrix W (size $t \times v$), where N is the number of EEG samples, t is the number of features for the brain-imaging data, and v is the number of the dimensions for the word vectors.

The data matrix $X^{N\times t}$ is then divided into training and testing subsets X_{train} and X_{test}. The word vector representations Y is similarly divided into Y_{train} and Y_{test}. The training subset X_{train} along with Y_{train} is then used to train multiple ridge regression models to learn the mapping from the brain imaging data (X_{train}) to the word vectors (Y_{train}) (see Figure 3.6). For our test set X_{test}, we average all the trials for a stimulus word within an age group resulting in 16 total samples (one sample of each stimulus word) to increase the signal-to-noise ratio. For example, if the test set contains five EEG samples for the word 'baby', we average these five samples to obtain one sample. This process is carried out for each word in the stimuli set, giving us 16 EEG samples for X_{test}.

During our analyses, we also tested the pre-onset decoding accuracy, which served as a baseline measurement. Thus, we use 1200ms of EEG data for each trial, ranging from -200ms to 1000ms, where 0ms marked the word stimuli onset.

We used Monte-Carlo nested cross-validation procedure with negative mean squared error to optimize the hyperparameter λ for the ridge regression model (see eq. 3.3). The Monte-Carlo procedure is useful when there are few data samples because it

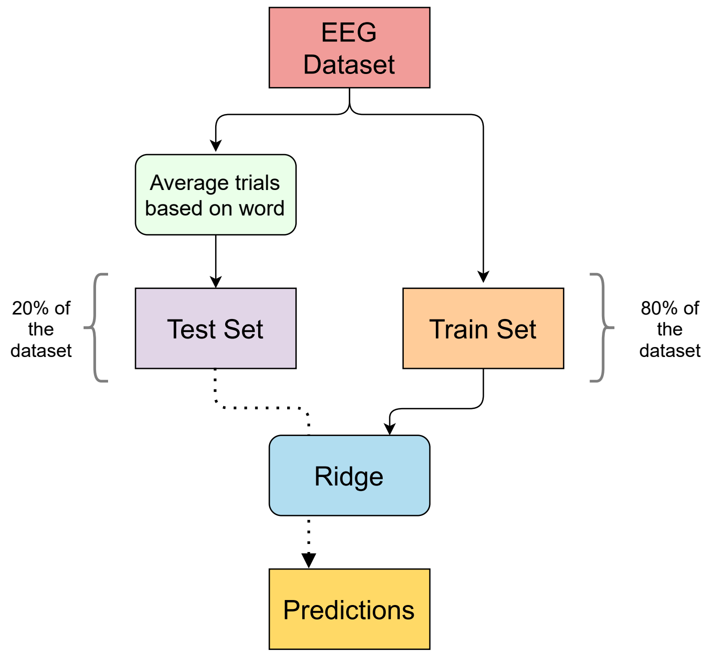


Figure 3.6: Train and Test sets. Randomly, 80% of the data is used as the train set and the remaining 20% of the data is used as the test set (train and test sets are mutually exclusive).

allows for a large number of train-test partitions to obtain the model performance estimate. During each sampling iteration of the Monte-Carlo procedure, we randomly sampled 80% of the dataset which constituted as X_{train} and the rest 20% constituted as the test set. For the inner loop of the nested cross-validation procedure, we used a k-fold split of the training set (X_{train}) with k=5 resulting in 20% of X_{train} being used as validation set for hyperparameter tuning. For the EEG data collected from 9-month-old infants (number of samples=1026), this resulted in 821 samples for X_{train} and 205 samples for X_{test} for each sampling iteration. About 164 samples from X_{train} were used for selecting the hyperparameter inside each cross-validation fold. For the EEG collected from 12-month-old infants, we had a total of 683 samples from which about 546 samples comprised of the training set (X_{train}), and the remaining samples comprised of X_{test}. About 109 samples from X_{train} were used for hyperparameter tuning. We averaged all the EEG samples in X_{test} to obtain 16 total samples (one sample for each word in the stimuli set). We used 50 Monte-Carlo sampling iterations for our analyses.

3.4.3 2 versus 2 Test

To evaluate the model performance, we used the 2 vs 2 test procedure. The test procedure works as follows. We first consider two instances from the predictions ($\hat{y}_i$, $\hat{y}_j$) and two instances from the ground truth vectors (y_i, y_j). We then use a distance metric to calculate the distance between the two pairs of vectors. We use the cosine distance criterion $d(u, v)$ where u and v are the word vectors. The chance accuracy of the 2 vs 2 test is 50% because there are two possible assignments of the predicted vectors to the ground truth vectors.

The 2 vs 2 test passes if:

$$d(y_i, \hat{y}_i) + d(y_j, \hat{y}_j) < d(y_i, \hat{y}_j) + d(y_j, \hat{y}_i) \quad (3.4)$$

The 2 vs 2 test is depicted in Figure 3.7. Considering two pairs of true labels and predicted labels, we calculate the cosine distance between the matching pairs of the true and predicted labels (green lines), and also calculate the cosine distance between the non-matching pairs of the true and predicted labels (red dashed lines). Then we use equation 3.4 to evaluate the 2 vs 2 test. The true and predicted labels are the vector representations of the word stimuli.

Since our test set contains the average of all the trials for a stimulus word, the total number of instances in our test set is 16. And the total number of pairs evaluated by the 2 vs 2 test are $\binom{16}{2} = 120$.

3.4.4 Testing for Above Chance Accuracy

To determine the analysis windows with above chance accuracy, we compared our results to random chance accuracy using the permutation test [36]. The permutation test is performed by randomizing the assignment of stimuli word vectors to their corresponding EEG data so that each word vector is randomly matched to any sample

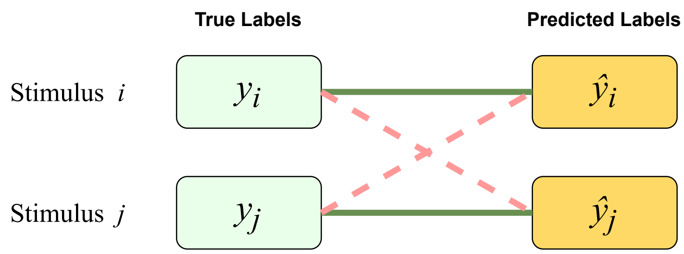


Figure 3.7: 2 vs 2 test showing pairwise comparisons of true and predicted labels. The test passes if the sum of the cosine distance between matching word vector pairs (represented by solid green lines) is less than the sum of the cosine distance between the non-matching word vector pairs (represented by red dashed lines).

of the EEG data. This mimics a situation where no association exists between the input EEG data X and the output word vectors Y. We ran our decoding framework on the dataset for 100 permutation iterations. As expected, since there is no relation between the input EEG data and the output word vectors, the accuracy when running the prediction framework on this shuffled dataset is chance (50% for 2 vs 2 test). We then fit a gaussian kernel density function to the obtained 2 vs 2 accuracies from the permuted dataset to generate a null distribution. Finally, we compared the non-permuted accuracies to the accuracies obtained from the permutation test to establish statistical significance.

Once the null distribution is obtained, a p-value is calculated for the original non-permuted accuracy. To correct the p-values for multiple comparisons over time ($\alpha < 0.01$), we use the Benjamini-Hochberg-Yekutieli False Discovery Rate (FDR) correction method with no dependency assumption [37], as it is a less conservative metric than other methods (e.g., Bonferroni correction) but still provides a robust for correcting for multiple comparisons.

3.4.5 Testing for Difference Between Conditions

After we trained our machine learning model to decode word semantics from EEG data, we used a different significance testing method called the non-parametric cluster permutation test [38] to compare the significant difference in accuracy values between different experimental conditions. The two conditions in our analyses are the two age groups (9-month-old and 12-month-old infants). The test is conducted as follows. First, we obtain the 2 vs 2 accuracies for each condition, resulting in fifty accuracy values for each sampling iteration (see 3.4.2). Next, we calculate a test statistic between the accuracy values for the two conditions, which gives us the observed test statistic. We then permute these accuracies by putting them into a single set, from which two random subsets of accuracies are drawn; this is called a random partition. We repeat this process 10000 times. We calculate the test statistic between the two subsets in each random partition and compare it to the observed test statistic, resulting in a p-value. Finally, we report the largest cluster of time points containing at least three or more consecutive windows with a significant difference in accuracy values ($p < 0.01$).

3.5 Conclusion

In this chapter, we discussed the various components of our study. A total of thirty-five infants participated in an audio-visual task while their EEG data was recorded. We then trained a simple machine learning model on the EEG data to predict the vector representations of the word stimuli. To evaluate model performance, we used the 2 vs 2 test, and the non-parametric cluster permutation test for finding significant differences between experimental conditions. In the next chapter, we will discuss the analyses conducted for tracking the neural representation of word semantics and discuss the results.

Chapter 4

Results and Discussion

4.1 Introduction

In chapter 3, we discussed the methodology of our study. In this chapter, we describe the various analyses to track the evolution of the semantic representation of words in infants. For each analysis, we first explain the steps involved and then observe the results. We start by inspecting individual word stimuli features such as animacy in Analysis 4.2 then move on to Analysis 4.3, where we discuss decoding overall semantic representation of words from neural responses. We also decode the phonetic representations of word stimuli in Analysis 4.4 and then inspect the similarity of the neural responses between the two age groups in Analysis 4.5. Lastly, we discuss the idea of fine-tuning word vectors for possibly better semantic modeling of infant neural data in Analysis 4.6. In the end, we summarize our results and discuss important findings.

4.2 Analysis 1: Decoding representations of Word Animacy

In the first experiment, we explored whether infants' neural representations include word animacy information. Since our stimuli set consists of 16 words equally divided into animate and inanimate words (see Chapter 3, Section 3.2.1 for details), we trained a classification model on the EEG data to predict whether a word belonged to the animate category or the inanimate category. For this binary classification problem, we trained a logistic regression model from scikit-learn [35] that takes the input EEG data X to predict the animacy category of the word Y. Unlike other analyses (described later in this chapter), the output is not the vector representation of the word stimuli but simply a binary label (0 or 1). We ran this analysis for each age group separately, and we trained and tested the model on the same age group.

To produce each accuracy point on the graph, we trained the model on windows of size 100ms with a 10ms sliding step giving us a timeseries of accuracy values. For each time window, we denote the accuracy at the end of the time window (for window 100-200ms, accuracy is shown at 200ms); this configuration is used for all the graphs in the rest of this thesis. Using 100ms windows will provide us with enough information from the EEG data to analyze the word stimuli. The animacy decoding results are shown in Figure 4.1 with significantly above chance accuracy ($p < 0.01$) and chance accuracy being 50%.

In Figure 4.1 results, we observed multiple points with above chance accuracy using EEG data collected from 9-month-old infants indicating that word animacy information is present in the neural responses of 9-month-old infants. After the word onset, the accuracy started rising at about 100ms and remained high until the end of the time course. The accuracy dipped quickly at around 350ms, after which it rose sharply at 400ms and stayed consistent till the end of the recording. On the other hand, when we attempted to decode the word animacy category from EEG data collected from 12-month-old infants, we did not observe above chance decoding accuracies anywhere in the time course. However, this does not necessarily indicate that animacy information is not encoded in the 12-month-old infants' neural responses. It is not exactly clear why word animacy cannot be decoded from 12-month-old infants' EEG data. It might be a case where the logistic regression model cannot learn the mapping between the EEG data and animacy category of the word, or it might just be a case of poor signal-to-noise ratio for decoding animacy, possibly due to the presence of EEG artefacts (movement, poor electrode conductance etc.). Further research is required to evaluate other machine learning models of binary classification for predicting word animacy from the EEG data.

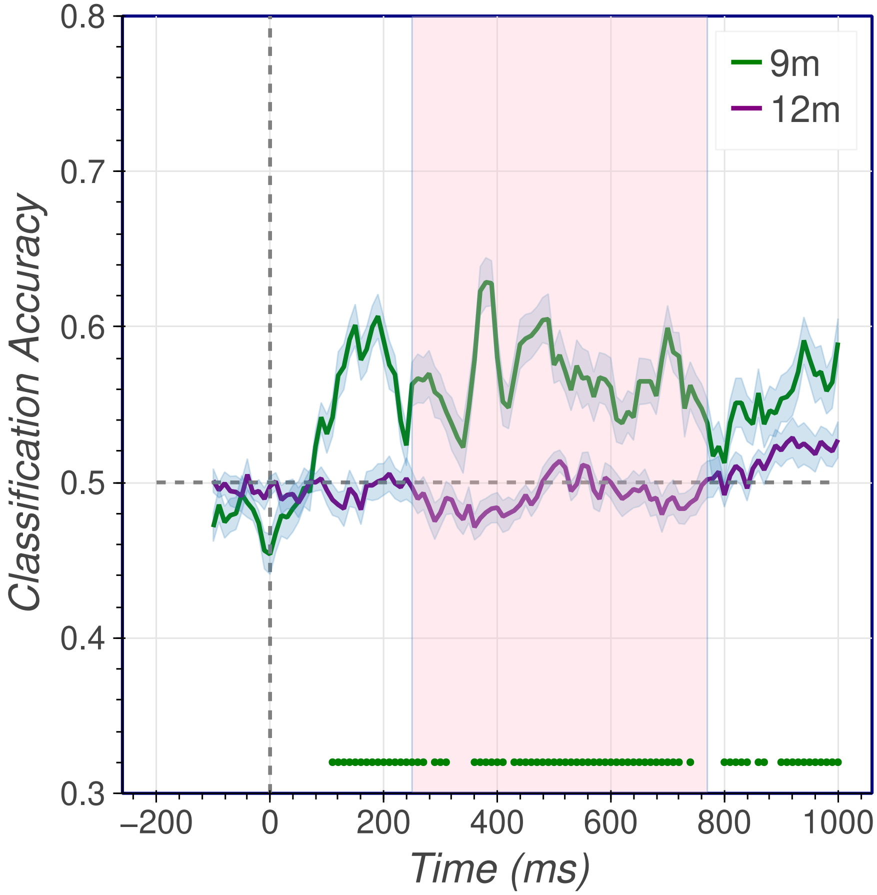


Figure 4.1: Accuracy for decoding word animacy category from the stimuli word. Green curve - 9-month-old infants; Purple curve - 12-month-old infants. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms to the left of the accuracy point) with a 10ms sliding step. Green dots show significantly above chance accuracy points for 9-month-olds (p-value < 0.01, FDR corrected for multiple comparisons over time). No significantly above chance accuracy points were obtained for 12-month-infants. The shaded (pink) area shows the significant difference in accuracies between the two age groups using non-parametric cluster permutation test [38].

4.3 Analysis 2: Representations of Word Semantics

In the previous section, we described the animacy classification analysis, where we used a simple logistic regression model to predict the animacy category of the word stimuli from the EEG data. In this section, we decode the overall semantic representation of words from the EEG data. To achieve this, we trained a ridge regression model on EEG data X, collected from infants to predict the 300-dimensional word vectors Y obtained from the pretrained Word2Vec model (see Chapter 3 Section 3.4.2 for details). Since these word vectors represent word semantics, significantly above chance accuracy will suggest the presence of word semantic information in the infant neural recordings. As we use the 2 vs 2 accuracy (see section 3.4.3) to evaluate model performance, the chance accuracy is 50%. We report significantly above chance accuracy $p < 0.01$. The results for the semantic decoding analysis is shown in Figure 4.2.

Referring to Figure 4.2, we found that we were able to decode the stimuli word vectors from the EEG data collected from 9-month-old and 12-month-old infants. This provides evidence of word semantic information in infant neural data recorded from both age groups. Since our analyses are carried out over 100ms chunks of EEG data with a 10ms sliding window, we can track the neural decoding results over time. Interestingly, for 9-month-old infants, just after the onset of the word stimulus, starting at around 60ms, all accuracy values were significantly above chance and remained high till the end of recording (green dots in Figure 4.2). This result suggests that infants as young as 9-month-old can distinguish between words in a basic vocabulary. We also observed two peaks of accuracy greater than 60% indicating two points in the trial (one at around 100ms and the second around 700ms), indicating that the word semantic representations in the neural recordings were the strongest at these two points. On the other hand, when the prediction framework was run on the

EEG data recorded from 12-months-old infants, we also found multiple above chance accuracy points (purple dots in Figure 4.2). The accuracy peaked close to 60% around 100ms and is sustained till about 450ms, after which it slowly decreased. However, the decodability of word vectors lasted only till 750ms, quickly dropping after that. We also observed that from around 620ms to 1000ms, there is a significant difference in accuracy values between the age groups (pink shaded area). This suggests that the neural representations for both groups of infants contain information about word semantics, but the neural representations of words are stronger for the 9-month-old infants than those of 12-month-old infants for the latter half of the trial. Such a result may indicate that 9-month-old infants hold on to the word semantic information longer than the 12-month-old infants to understand the word stimuli. In some research that uses looking time paradigms, older infants are quicker in directing their gaze to the matching picture of a spoken word than younger infants [1]. This may support the hypothesis that younger infants need longer to understand the meaning of a spoken word. However, more research is required to investigate such an effect.

In this analysis, we saw that the evoked neural responses in infants from both age groups were correlated with the stimuli word vector representations. However, we were unsure whether the above chance accuracy could be attributed to the word semantics only there other aspects that contributed to such a high decoding accuracy. We investigate this issue in the next analysis.

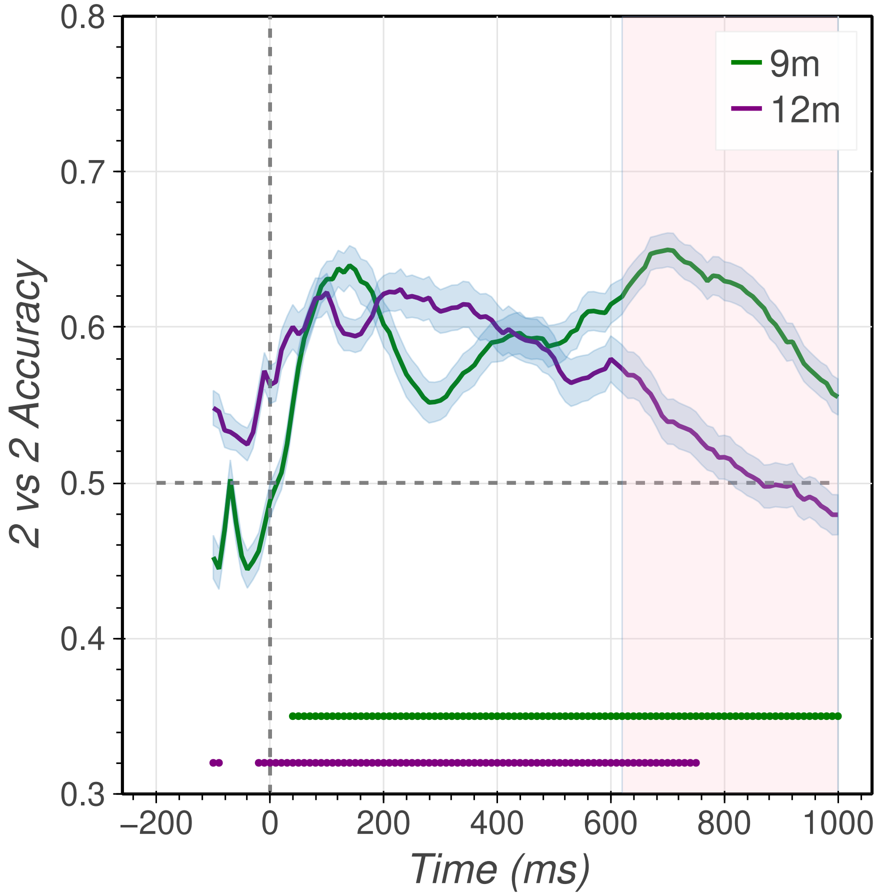


Figure 4.2: 2 vs 2 accuracy for predicting pretrained Word2Vec word vectors from EEG data collected from 9-month-old and 12-month-old infants. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms to the left of the accuracy point) with a 10ms sliding step. The green dots above the x-axis represent the points where the accuracy is reliable for 9-month-old infants, and the purple dots represent reliable above chance accuracy for 12-month-old infants (p-value < 0.01, FDR corrected for multiple comparisons over time).

4.4 Analysis 3: Presence of Phonetic representations of Word Stimuli

In Analysis 4.3, we decoded the overall word semantic representations from the EEG data. We were also interested in the presence of other aspects of word stimuli in the EEG data. As we know that infants in early age groups are very sensitive to phonetic differences in speech sounds [2, 39], we investigated the presence of word phonetics in the neural responses.

We carry out this analysis as follows. Each word in our stimuli set comprises multiple letters (ranging between three to six letters). We map the letters to their respective phonemes, where each phoneme was used in its IPA notation[1]. Whenever possible, we used the Canadian pronunciation, and the General American (GA) pronunciation if the Canadian pronunciation was not available, defined in Dictionary.com[2] and Wiktionary[3]. Then, we map each phoneme to a 36-dimensional vector obtained from [40]. We concatenated each phoneme vector into one long vector P and zero-padded all the concatenated vectors to make them of equal size, resulting in a 216-dimensional long vector. Finally, we predict these 216-dimensional vectors from the EEG data using a ridge regression model. The process of creating these concatenated phoneme vectors is shown in 4.3.

The presence of phonetic components in the neural responses will be indicated by significantly above chance accuracy. Figure 4.4 shows the results of predicting the phoneme vectors from the neural responses.

It is interesting to see that for 9-month-old infants (orange curve in Figure 4.4a, we were able to decode phonetic information from EEG data with significant above chance accuracy. This indicates that phonetic information is encoded in the neural responses of 9-month-old infants. We observed significantly above chance accuracy immediately after word onset, which lasted until 1000ms. On the other hand, when we observe the accuracy for the 12-month-old infants (black curve in Figure 4.4a), we notice an initial rise in decoding accuracy around word onset, which lasts till around 380ms and quickly dropping after that. This suggests that even though the EEG data from 12-month-old infants encode phonetic information, it is transient and not as long lasting as the phonetic representations for the 9-month-old infants. The pink shaded area shows regions with significant differences between decoding accuracy for phoneme vectors for 9-month-old infants and 12-month-old infants. Starting around 400ms, the

[1] https://en.wikipedia.org/wiki/International_Phonetic_Alphabet
[2] https://www.dictionary.com/
[3] https://www.wiktionary.org/

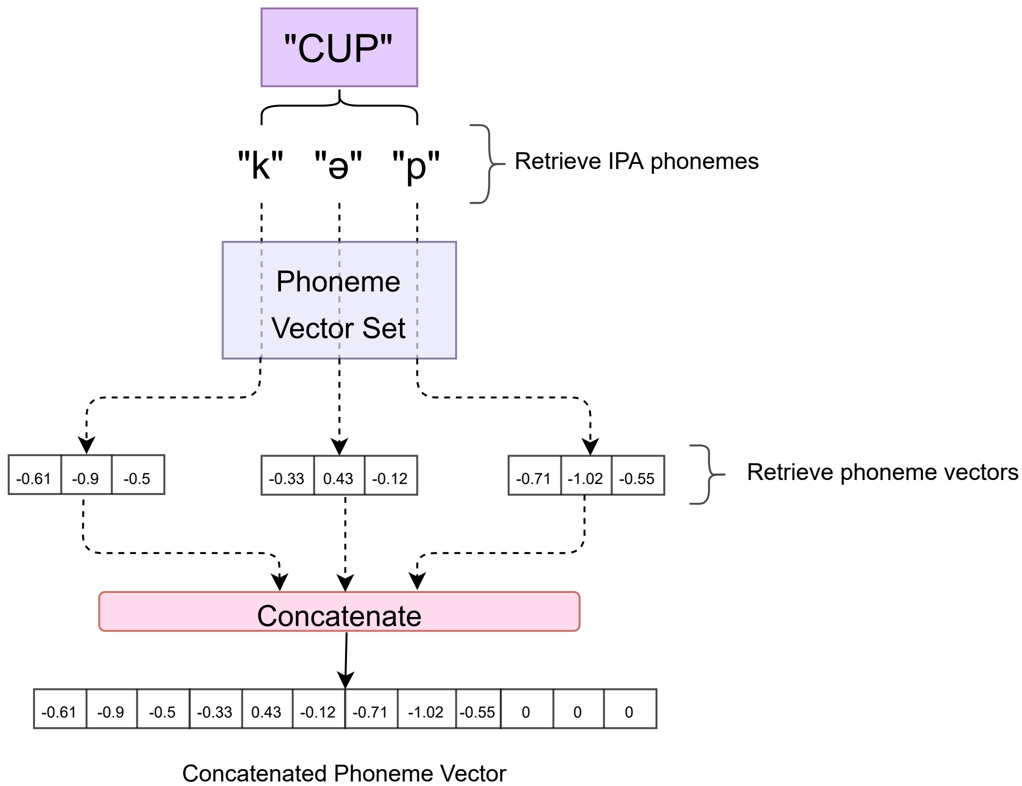


Figure 4.3: Phoneme vector creation process for a sample word *cup* from the stimuli set. The stimulus word is broken down into its individual IPA phonemes. For each IPA phoneme, a vector is retrieved from Mielke *et al.* [40]. Finally, a long vector is created by concatenating individual phoneme vectors. For shorter words, zero-padding is used to obtain equal length vectors. We show 3-dimensional phoneme vectors for simplicity. Actual vectors are 36 dimensions long.

neural representations differ greatly, resulting in a divergence of the accuracy curves.

We also compare the results between the current phoneme decoding analysis and the semantic decoding analysis described in Analysis 4.3, in Figure 4.4b for 9-month-old infants and in Figure 4.4c for 12-month-old infants.

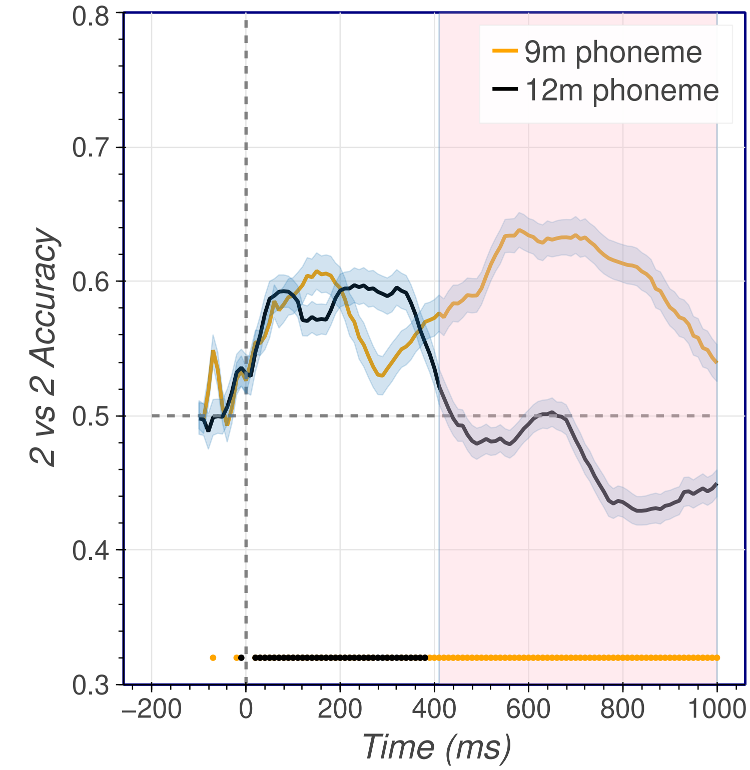
In Figure 4.4b, we see that for 9-month-old infants, the accuracy curve for the phoneme decoding analysis (orange) traces close to the accuracy curve for the semantic decoding analysis (green). On the flip side, when we compare the phoneme decoding analysis and semantic decoding analysis for 12-month-old infants, we see that in Figure 4.4c, the curve for phoneme decoding analysis (black) and semantic decoding analysis (purple) trace each other closely till around 400ms after which the phoneme decoding accuracy drops sharply but semantic decoding accuracy persists till about 750ms. This effect is depicted by the pink shaded area showing a significant difference in accuracy from around 370ms to 1000ms.

In this analysis, we tried to understand the presence of word phonetic information in the neural responses. We observed that the EEG data for infants in both age groups captures the phonetic representations of words, but the phonetic decoding accuracy in the two age groups differs significantly. The next section will investigate how similar the neural responses are across the two age groups for our stimuli set.

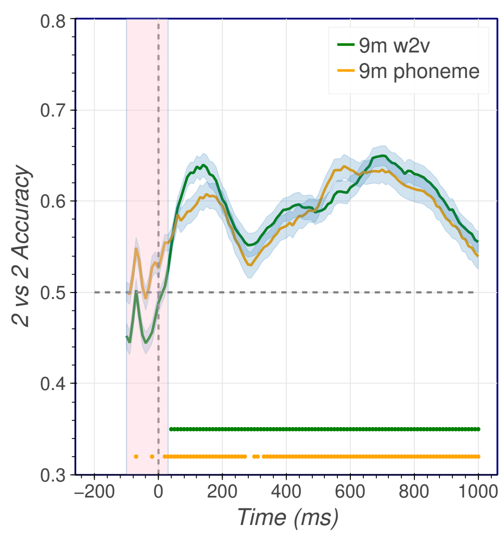
4.5 Analysis 4: Shared Representations of Word Semantics between age groups

All the previous experiments trained and evaluated machine learning models on the same age group, which investigated the decodability of word stimuli from the neural responses. We were also interested in the similarity of neural responses for the stimuli between age groups. In other words, are the neural responses from one age group similar to the other age group? We can answer this question by running the prediction framework on EEG data recorded from one age group (for example, 9-month-old infants) and evaluate its performance on the EEG data collected from the other age group (for example, 12-month-old infants). Suppose the pattern learned by the machine learning model from one age group is useful for decoding the word semantics from the EEG data obtained from the other age group. In that case, it will indicate that the neural responses are similar across the two age groups.

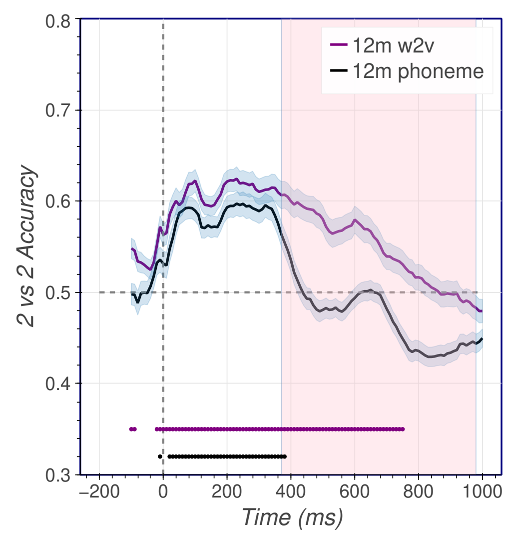
Similar to Analysis 4.3, we train the model on the EEG data to predict the 300-dimensional word vectors obtained from the Word2Vec model. Successful decoding across age groups will be indicated by significantly above chance accuracy.



(a) Decoding accuracy for predicting phonemes from EEG data recorded from 9-month-old and 12-month-old infants.



(b) Comparison of accuracy for predicting phonemes (orange) and Word2Vec semantic vectors (green) from EEG data recorded from 9-month-old infants.



(c) Comparison of accuracy for predicting phonemes (black) and Word2Vec semantic vectors (purple) from EEG data recorded from 12-month-old infants.

Figure 4.4: Phoneme decoding accuracy and its comparison to the decoding accuracy of pretrained Word2Vec word vectors. Shaded areas show significant differences in accuracies, and dots above the x-axis show points of significantly above chance accuracy (p-value < 0.01).

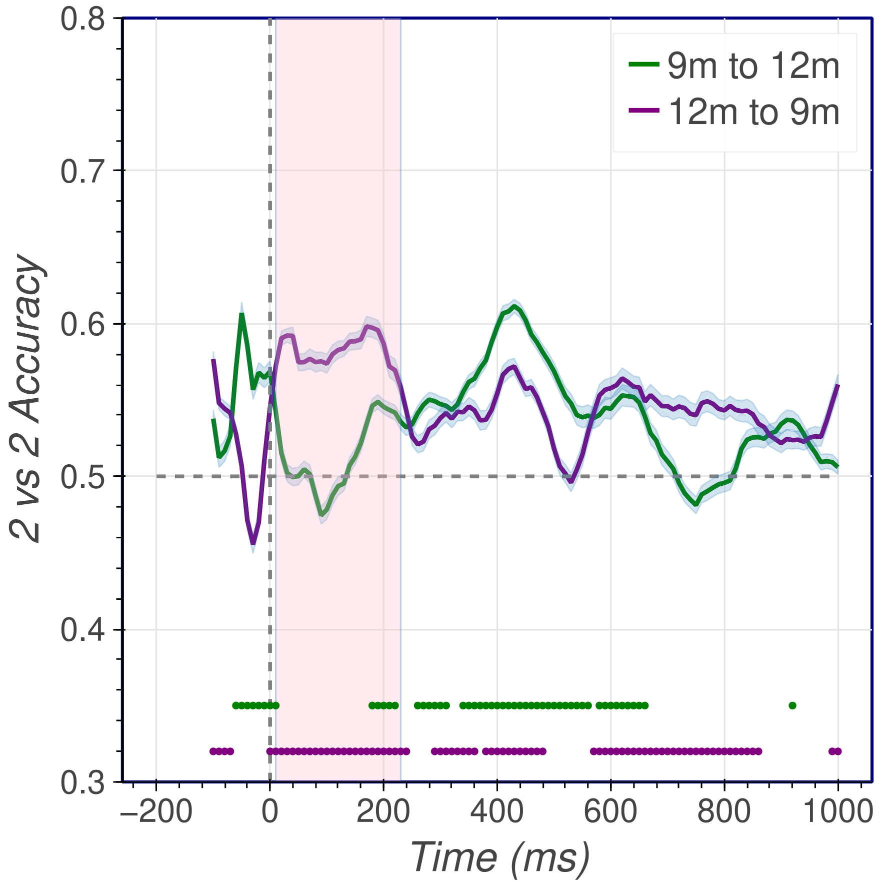


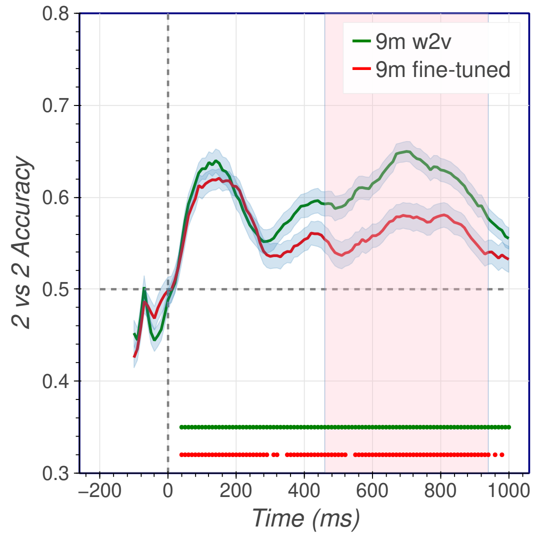
Figure 4.5: 2 vs 2 accuracy for predicting pretrained Word2Vec word vectors from EEG data across age groups. Each point on the graph represents an accuracy value for the model trained on a 100ms window (100ms to the left of the accuracy point) with a 10ms sliding step. The green dots denote above chance accuracy for the model trained on 9-month-old and tested on 12-month-old infants, and purple dots denote above chance accuracy for the model trained on 12-month-old infants and tested on 9-month-old infants (p-value < 0.01, FDR corrected for multiple comparisons over time). The shaded area shows where the accuracy curve is significantly different for the two conditions.

In Figure 4.5, when we trained the machine learning model on 9-month-old infants' EEG and tested on 12-month-old infants' EEG data (green curve), we observed significantly above chance accuracy. Similarly, when we trained the model on 12-month-old infants' EEG data, we predicted the word vectors from 9-month-old infants' EEG data with significantly above chance accuracy (purple curve). This suggests that neural representations for both age groups are similar at multiple time points.

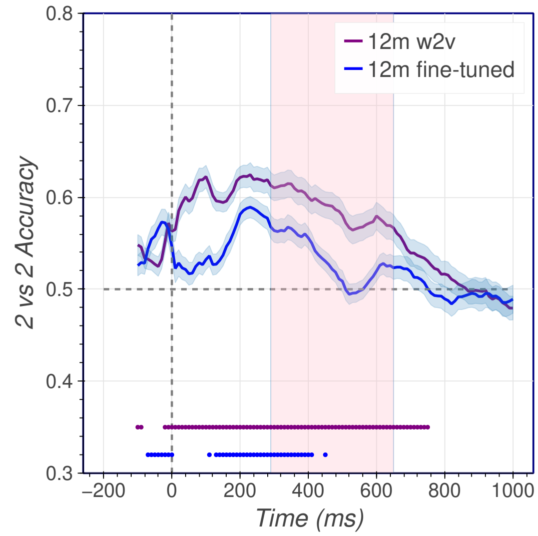
4.6 Analysis 5: Decoding Fine-tuned Representations of Word Semantics

In Analysis 4.3, we ran our prediction framework to predict the word vectors obtained from the pretrained Word2Vec model. Since the pretrained Word2Vec model is trained on the Google News Corpus, the word vectors will likely be oriented to the semantic representations of adults. Perhaps, we can change these word vectors to better represent the semantic representation of words in infants. Word vectors characteristic of infants' representation of word semantics may lead to better infant neural data predictions than pretrained word vectors. To this end, we fine-tuned the pretrained Word2Vec model on two datasets, the children's book test dataset [41] and the Childes corpus dataset [42]. We used the Word2Vec implementation in the Gensim Python Package [43], which we fine-tuned using one thousand iterations of gradient descent with default learning rates; $start = 0.025$ and reduces to $end = 0.0001$ with the iterations of gradient descent. Finally, we proceeded to decode these fine-tuned word vector representations from the EEG data for which the results are shown in 4.6.

In Figure 4.6 we see that fine-tuned word vectors can be decoded with significantly above chance accuracy using the EEG data recorded from both groups of infants. However, we see a clear difference when comparing the results with the decoding accuracy of pretrained (non fine-tuned) word vectors. In Figure 4.6a, we notice that the decoding accuracy for pretrained word vectors (green) is high in the later stages of the time course compared to the decoding accuracy for the fine-tuned word vectors (orange). The pink shaded area shows the region where the accuracy values are significantly different. Similarly, for 12-month-old infants, we see that in Figure 4.6b the decoding accuracy for pretrained word vectors (purple) is generally higher than the accuracy of fine-tuned word vectors (blue).



(a) Accuracy comparison for predicting fine-tuned word vectors (orange) and pretrained word vectors (green) from EEG data collected from 9-month-old infants.



(b) Accuracy comparison for predicting fine-tuned word vectors (blue) and pretrained word vectors (purple) from EEG data collected from 12-month-old infants.

Figure 4.6: Decoding accuracy for predicting fine-tuned word vectors and its comparison to pretrained Word2Vec word vector decoding accuracy. Shaded areas show significant differences in accuracies, and dots above the x-axis show points of significantly above chance accuracy (p-value < 0.01).

4.7 Discussion

Our results show that we can decode word stimuli from EEG data recorded from infants using a machine learning approach. First, we showed that word animacy can be decoded from neural responses. Second, we showed that overall word semantics can be decoded from the neural responses from both groups of infants. Third, we observed that the neural responses contained strong representations of the phonetic information of the word stimuli. Fourth, we observed that there exist similar representations of word semantics across 9-month-old and 12-month-old infants. Finally, we explored a fine-tuning approach to make more 'child-like' word semantic representations.

In the animacy decoding analysis (Analysis 4.2), we observed various points of significantly above chance decoding accuracy from 9-month-old infants' EEG data, suggesting the presence of strong representations of word animacy in the neural responses. However, no significantly above chance accuracies were observed for animacy decodability for EEG data recorded from 12-month-old infants.

This inability to decode the word animacy category does not necessarily mean 12-month-old infants do not have a representation of word animacy. It may just be the case that the Logistic Regression model is not able to find patterns of word animacy in the EEG data from 12-month-old infants, possibly due to EEG movement artefacts or bad signal-to-noise ratio for word animacy. In addition, it also does not imply that 12-month-old's EEG data cannot be used to decode word semantics, as we noticed in Analysis 4.3.

While decoding the semantic word vectors in Analysis 4.3 (Figure 4.2), we found that the neural responses of both groups of infants contained strong representations of word semantics. In addition to previous studies [2, 3], which showed that infants could detect semantic mismatch of the stimuli, the presence of multiple above chance accuracy points in this analysis shows that infants indeed have a semantic representation of the stimuli word as well. We also observed a significant difference in decoding accuracy in the latter half of the trial for the two groups of infants (pink shaded area in Figure 4.2). However, more research is required to investigate such a difference in accuracy.

Although we observed that overall semantic information can be decoded from the EEG data of both groups of infants, we were also interested in decoding other aspects of the word stimuli. In Analysis 4.4, we discovered that phonetic information of the stimuli words can also be decoded from both groups of infants with significantly above chance accuracy.

We observed that the phonetic decoding accuracy for 9-month-old infants closely traced the decoding accuracy curve for semantics (Figure 4.4b). This might suggest that a majority of neural signal used for decoding word semantics from 9-month-old infants is correlated with the phonetic components of the words. On the other hand, for 12-month-old infants, the phonetic information was decodable starting after word onset. However, it lasted only till around 400ms, unlike word semantic decodability, which lasted till around 750ms (Figure 4.4c), suggesting that the neural representations for 12-month-old infants in the latter half of the trial are likely correlated to word semantics rather than word phonetics. Observing both, Analysis 4.3 and Analysis 4.4, we cannot rule out the possibility that the decodability of word phonemes may act as a confound to the semantic decoding accuracy for both age groups. However, for 9-month-old infants, the effect exists for the whole recording, while lasting only till around 400ms for the 12-month-old infants.

We also investigated the similarity of the neural representations across age groups. In Analysis 4.5, we observed that we were able to predict the word semantic representation from either age group, indicating that the mental representations of words are similar (Figure 4.5).

Such a result also suggests that even though developmental changes in neural responses between the infants of the two age groups might exist, there is an overwhelming similarity between them.

In the final analysis (Analysis 4.6), we explored a fine-tuning approach to obtain word vectors that are more characteristic of the semantic representations of infants. Surprisingly, we did not obtain better decoding accuracy using fine-tuned vectors, which might have been caused due to the small size of the dataset and the non-uniform distribution of word tokens. Nevertheless, we believe that improved word vector representations of semantics could be developed for infants with datasets that are larger, containing more occurrences of the word stimuli, which may lead to better word vectors for representing word semantics in infants, possibly resulting in better decoding accuracy.

4.8 Conclusion

This chapter described the various analyses that decode the infant mind and scan for word semantic information. We started by describing the analysis where we searched for word animacy from the brain-imaging data. Next, we found the presence of semantic information of words in the brain imaging data by accurately predicting word vectors. We also observed significantly above chance accuracy for decoding phonetic components of the word stimuli for both groups of infants. Next, we discovered that neural responses for the two age groups contain similar information for the word stimuli. Finally, we discussed the process of fine-tuning word vectors as a possible approach for better modeling of infant semantic representation.

Chapter 5

Conclusion & Future Work

While a vast body of research exists for tracking changes in the neural representation of word semantics in adults [5, 6, 9, 12], there is a lack of research that examines the neural representation of word semantics in early infants. We try to reduce this research gap by using a decoding approach to predict word vectors from brain activity recorded from infants. Using the techniques described in this thesis, we examined the changes in neural patterns as the infants heard single word stimuli.

First, we summarised studies that have successfully utilized semantic models of text and brain-imaging techniques to analyze the semantics representation of words in the brain. In addition, we also discussed recent studies that have found the presence of word semantics in infants. For our work, we used EEG data recorded from infants as they heard single word stimuli. We then observed the evolution of the neural responses over time. We first investigated the presence of basic semantic properties such as animacy, after which we proceeded to decode semantic information from the brain. We also inspected the presence of word phonetics in the infants' neural responses. Finally, we explored the similarity of neural representations of words in the two age groups.

Summarizing the contributions, we found that,

- A machine learning approach is useful for investigating the semantic representation of words from 9-month-old and 12-month-old infants.

- Semantic information can be decoded from both 9-month-old and 12-month-old infants immediately after word onset.
- Individual word stimuli properties such as phonetic information can be decoded from neural recordings from both groups of infants.
- Neural representation of word semantics are similar for both age groups.

Besides conventional techniques such as ERP, we provided evidence that machine learning techniques effectively decode word representations from infant brain-imaging data.

Even though we showed that word semantics can be decoded from infant brain activity, this work only decodes single word semantics. A natural extension would be to explore the decodability of more complex linguistic elements such as phrases from infant brain activity which we hope to explore in the future.

Bibliography

[1] E. Bergelson and D. Swingley, "At 6-9 months, human infants know the meanings of many common nouns," *Proceedings of the National Academy of Sciences*, vol. 109, no. 9, pp. 3253–3258, 2012. DOI: 10.1073/pnas.1113380109.

[2] E. Parise and G. Csibra, "Electrophysiological Evidence for the Understanding of Maternal Speech by 9-Month-Old Infants," *Psychological science*, vol. 23, no. 7, pp. 728–733, Jul. 2012, ISSN: 0956-7976. DOI: 10.1177/0956797612438734. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4641316/ (visited on 06/20/2021).

[3] M. Friedrich and A. D. Friederici, "Maturing brain mechanisms and developing behavioral language skills," eng, *Brain and Language*, vol. 114, no. 2, pp. 66–71, Aug. 2010, ISSN: 1090-2155. DOI: 10.1016/j.bandl.2009.07.004.

[4] R. Saha, J. Campbell, J. F. Werker, and A. Fyshe, "Insights into early word comprehension: Tracking the neural representations of word semantics in infants," 2021. DOI: 10.1101/2021.10.28.466334.

[5] A. Fyshe, G. Sudre, L. Wehbe, N. Rafidi, and T. M. Mitchell, "The lexical semantics of adjective–noun phrases in the human brain," *Human Brain Mapping*, vol. 40, no. 15, pp. 4457–4469, 2019. DOI: 10.1002/hbm.24714.

[6] C. Foster, C. C. Williams, O. E. Krigolson, and A. Fyshe, "Using EEG to decode semantics during an artificial language learning task," *Brain and Behavior*, 2021. DOI: 10.1002/brb3.2234.

[7] B. Murphy, M. Poesio, F. Bovolo, L. Bruzzone, M. Dalponte, and H. Lakany, "EEG decoding of semantic category reveals distributed representations for single concepts," en, *Brain and Language*, vol. 117, no. 1, pp. 12–22, Apr. 2011, ISSN: 0093-934X. DOI: 10.1016/j.bandl.2010.09.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0093934X10001811 (visited on 05/30/2021).

[8] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with the meanings of nouns," eng, *Science (New York, N.Y.)*, vol. 320, no. 5880, pp. 1191–1195, May 2008, ISSN: 1095-9203. DOI: 10.1126/science.1152876. [Online]. Available: http://www.sciencemag.org/cgi/content/full/320/5880/1191 (visited on 09/05/2021).

[9] G. Sudre, D. Pomerleau, M. Palatucci, L. Wehbe, A. Fyshe, R. Salmelin, and T. Mitchell, "Tracking neural coding of perceptual and semantic features of concrete nouns," eng, *NeuroImage*, vol. 62, no. 1, pp. 451–463, Aug. 2012, ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2012.04.048.

[10] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic Regularities in Continuous Space Word Representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia: Association for Computational Linguistics, Jun. 2013, pp. 746–751. [Online]. Available: https://aclanthology.org/N13-1090 (visited on 09/25/2021).

[11] A. Papadimitriou, N. Passalis, and A. Tefas, "Decoding Generic Visual Representations From Human Brain Activity using Machine Learning," *arXiv:1811.01757 [cs, q-bio]*, Nov. 2018, arXiv: 1811.01757. [Online]. Available: http://arxiv.org/abs/1811.01757 (visited on 04/07/2021).

[12] M. Honari-Jahromi, B. Chouinard, E. Blanco-Elorrieta, L. Pylkkänen, and A. Fyshe, "Neural representation of words within phrases: Temporal evolution of color-adjectives and object-nouns during simple composition," en, *PLOS ONE*, vol. 16, no. 3, e0242754, Mar. 2021, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0242754. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242754 (visited on 04/07/2021).

[13] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in Neuroinformatics*, vol. 8, p. 14, 2014, ISSN: 1662-5196. DOI: 10.3389/fninf.2014.00014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fninf.2014.00014 (visited on 09/05/2021).

[14] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: A tutorial overview," *NeuroImage*, vol. 45, no. 1 Suppl, S199–S209, Mar. 2009, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2008.11.007. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2892746/ (visited on 09/05/2021).

[15] T. Grootswagers, S. G. Wardle, and T. A. Carlson, "Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data," eng, *Journal of Cognitive Neuroscience*, vol. 29, no. 4, pp. 677–697, Apr. 2017, ISSN: 1530-8898. DOI: 10.1162/jocn_a_01068.

[16] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: Multi-voxel pattern analysis of fMRI data," eng, *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, Sep. 2006, ISSN: 1364-6613. DOI: 10.1016/j.tics.2006.07.005.

[17] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," *Communications of the ACM*, vol. 8, no. 10, pp. 627–633, 1965. DOI: 10.1145/365628.365657.

[18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Jan. 2013. arXiv: 1301.3781 `[cs.CL]`.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems*, vol. 26, Curran Associates, Inc., 2013.

[20] T. M. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman, "Learning to Decode Cognitive States from Brain Images," en, *Machine Learning*, vol. 57, no. 1, pp. 145–175, Oct. 2004, ISSN: 1573-0565. DOI: 10.1023/B:MACH.0000035475.85309.1b. [Online]. Available: https://doi.org/10.1023/B:MACH.0000035475.85309.1b (visited on 08/24/2021).

[21] A. Fyshe, P. P. Talukdar, B. Murphy, and T. M. Mitchell, "Interpretable semantic vectors from a joint model of brain and text based meaning," The 52nd Annual Meeting of the Association for Computational Linguistics , Baltimore, Maryland. 2014, 2014. [Online]. Available: https://www.cs.cmu.edu/~afyshe/papers/acl2014/jnnse_acl2014.pdf.

[22] S. Gibbon, A. Attaheri, Á. N. Choisdealbha, S. Rocha, P. Brusini, N. Mead, P. Boutris, H. Olawole-Scott, H. Ahmed, S. Flanagan, K. Mandke, M. Keshavarzi, and U. Goswami, "Machine learning accurately classifies neural responses to rhythmic speech vs. non-speech from 8-week-old infant EEG," *Brain and Language*, vol. 220, 2021.

[23] M. Zare, Z. Rezvani, and A. A. Benasich, "Automatic classification of 6-month-old infants at familial risk for language-based learning disorder using a support vector machine," en, *Clinical Neurophysiology*, vol. 127, no. 7, pp. 2695–2703, Jul. 2016, ISSN: 1388-2457. DOI: 10.1016/j.clinph.2016.03.025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245716300074 (visited on 08/21/2021).

[24] L. L. Emberson, B. D. Zinszer, R. D. S. Raizada, and R. N. Aslin, "Decoding the infant mind: Multivariate pattern analysis (MVPA) using fNIRS," en, *PLOS ONE*, vol. 12, no. 4, e0172500, Apr. 2017, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0172500. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172500 (visited on 08/07/2021).

[25] A. D. Friederici, "Neurophysiological markers of early language acquisition: From syllables to sentences," en, *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 481–488, Oct. 2005, ISSN: 1364-6613. DOI: 10.1016/j.tics.2005.08.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136466130500241X (visited on 09/06/2021).

[26] M. Kutas and K. D. Federmeier, "Electrophysiology reveals semantic memory use in language comprehension," en, *Trends in Cognitive Sciences*, vol. 4, no. 12, pp. 463–470, Dec. 2000, ISSN: 1364-6613. DOI: 10.1016/S1364-6613(00)01560-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661300015606 (visited on 09/06/2021).

[27] L. Fenson, *MacArthur Communicative Development Inventories: User's Guide and Technical Manual*, en. Cengage Learning, Jul. 1993, Google-Books-ID: A9iGKgAACAAJ, ISBN: 9781565932593.

[28] R. Tincoff and P. W. Jusczyk, "Some Beginnings of Word Comprehension in 6-Month-Olds," en, *Psychological Science*, vol. 10, no. 2, pp. 172–175, Mar. 1999, ISSN: 0956-7976. DOI: 10.1111/1467-9280.00127. [Online]. Available: https://doi.org/10.1111/1467-9280.00127 (visited on 07/05/2021).

[29] R. Tincoff and P. W. Jusczyk, "Six-month-olds comprehend words that refer to parts of the body," *Infancy*, vol. 17, no. 4, pp. 432–444, 2012, ISSN: 1532-7078(Electronic),1525-0008(Print). DOI: 10.1111/j.1532-7078.2011.00084.x.

[30] E. Bergelson and D. Swingley, "Early Word Comprehension in Infants: Replication and Extension," *Language learning and development : the official journal of the Society for Language Development*, vol. 11, no. 4, pp. 369–380, 2015, ISSN: 1547-5441. DOI: 10.1080/15475441.2014.979387. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4671511/ (visited on 06/14/2021).

[31] E. Bergelson and D. Swingley, "Young Infants' Word Comprehension Given an Unfamiliar Talker or Altered Pronunciations," *Child development*, vol. 89, no. 5, pp. 1567–1576, Sep. 2018, ISSN: 0009-3920. DOI: 10.1111/cdev.12888. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5741549/ (visited on 07/05/2021).

[32] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime: User's Guide. Reference Guide. Getting Started Guide*. Psychology Software Tools, Incorporated, 2002. [Online]. Available: http://www.pstnet.com/eprime.cfm.

[33] Y.-P. Ruan, Z.-H. Ling, and Y. Hu, "Exploring Semantic Representation in Brain Activity Using Word Embeddings," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 669–679. DOI: 10.18653/v1/D16-1064. [Online]. Available: https://www.aclweb.org/anthology/D16-1064 (visited on 06/03/2021).

[34] S. L. Kivisaari, M. van Vliet, A. Hultén, T. Lindh-Knuutila, A. Faisal, and R. Salmelin, "Reconstructing meaning from bits of information," en, *Nature Communications*, vol. 10, no. 1, p. 927, Feb. 2019, ISSN: 2041-1723. DOI: 10.1038/s41467-019-08848-0. [Online]. Available: https://www.nature.com/articles/s41467-019-08848-0 (visited on 06/03/2021).

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[36] M. Ojala and G. C. Garriga, "Permutation Tests for Studying Classifier Performance," *Journal of Machine Learning Research*, vol. 11, no. 62, pp. 1833–1863, 2010. [Online]. Available: http://jmlr.org/papers/v11/ojala10a.html (visited on 03/16/2021).

[37] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *The Annals of Statistics*, vol. 29, no. 4, pp. 1165–1188, Aug. 2001, ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1013699998. [Online]. Available: https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-4/The-control-of-the-false-discovery-rate-in-multiple-testing/10.1214/aos/1013699998.full (visited on 05/30/2021).

[38] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," eng, *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, Aug. 2007, ISSN: 0165-0270. DOI: 10.1016/j.jneumeth.2007.03.024.

[39] C. L. Stager and J. F. Werker, "Infants listen for more phonetic detail in speech perception than in word-learning tasks," *Nature*, vol. 388, no. 6640, pp. 381–382, 1997. DOI: 10.1038/41102.

[40] J. Mielke, "A phonetically based metric of sound similarity," en, *Lingua*, Phonological Similarity, vol. 122, no. 2, pp. 145–163, Jan. 2012, ISSN: 0024-3841. DOI: 10.1016/j.lingua.2011.04.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0024384111000891 (visited on 05/29/2021).

[41] F. Hill, A. Bordes, S. Chopra, and J. Weston, "The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations," *arXiv:1511.02301 [cs]*, Apr. 2016, arXiv: 1511.02301. [Online]. Available: http://arxiv.org/abs/1511.02301 (visited on 08/04/2021).

[42] B MacWhinney, "The childes project: Tools for analyzing talk. Third edition.," *Mahwah, NJ: Lawrence Erlbaum Associates*, 2000.

[43] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," English, in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, http://is.muni.cz/publication/884893/en, Valletta, Malta: ELRA, May 2010, pp. 45–50.