Children's Vocabulary Development in American Sign Language*

Similar Trends Between Developing Vocabulary Production in Sign Language and Spoken Language

Aliyah Maxine Ramos

20 April 2023

Abstract

This paper investigates the development of American Sign Language in children by observing the relationship between the range of ages during developmental years and the various types of vocabulary terms learned. Using the data from the open database, Wordbank, we found that children learning to speak a visual language share a similar trend in vocabulary development to those learning spoken languages. This highlights the importance of learning American Sign Language at an early age as it provides the ability for those with hearing impairments or disabilities to participate in society like any other able-bodied individual.

1 Introduction

American Sign Language (ASL) allows for communication and interaction with the community of people who are deaf or have hearing disabilities. This makes sign language an essential tool for inclusivity and accessibility. Learning ASL enables those with hearing impairments to participate in society just like any other abled person. The ability to communicate with others through ASL can help diminish communication barriers and help create more inclusive environments for all. And learning languages is best at early stages of life.

In this paper, we investigate the patterns and trends that occur during the early developmental years of childhood in terms of vocabulary in American Sign Language. The estimand that we are trying to estimate is the number of vocabulary words in ASL that a child can produce at a given developmental age. We aimed to visualize the milestones for children's vocabulary development in sign language to compare it to the timeline of vocabulary development for spoken languages. This paper uses data obtained from the open database, Wordbank,

*Code and data are available at: https://github.com/mxnrms/Children_Vocab_Development.git.

which exhibits information on the development of children's vocabulary in various languages. The dataset, "Full Child-by-Word", available on this database, provides information on the demographic details of each child as well as the variables presenting their vocabulary assessment.

With our focus on the development of American Sign Language vocabulary, we found that children learning to communicate a visual language share similar trends in vocabulary development to those that are learning a spoken language. We were also interested in how children develop this natural language at an early age and the differences in comprehension and production of vocabulary items between ages. The findings in this paper highlight the importance of learning ASL at an early age because it allows those with hearing disabilities to experience a more accessible world. The importance of learning ASL at an early age is emphasized as it provides the ability for those with hearing impairments or disabilities to participate in society like any other able-bodied individual. It acknowledges how the world often caters to able-bodied members of society, making things less accessible for the deaf and hard-of-hearing community. By understanding the cognitive processes involved in learning a visual language, we are better able to support individuals in their language acquisition.

The remainder of this paper is structured as follows: Section 2 introduces the data used for analysis and findings, including visualizations of the variables of interest, Section 3 presents the simple linear regression model to justify and predict the relationship between the developmental ages of children and their production of vocabulary words in ASL, Section 4 displays the interpretations of the model alongside other findings from analyzing the data, and Section 5 provides a discussion on the implications of the findings as well as the weaknesses of this paper and its next steps for further study on this subject.

2 Data

2.1 Data Source

The data utilized in this paper concentrates on the vocabulary growth of children, which is accessible in several languages through the open database, Wordbank (Frank et al. 2016b). At present, Wordbank contains data concerning children's vocabulary development in 29 different languages, covering age groups from as early as 12 months to 36 months. The data source, Wordbank, uses open-source tools to offer openly available data to the public, which can be downloaded for free from both its website and GitHub repository. The primary entities of the Wordbank repository consist of the instrument, which refers to the survey or questionnaire form, and the item, which refers to a specific question in the instrument, along with the administration, unique child, and the language.

The data collection process of the database includes parent reports, naturalistic observation, and experimental testing. The primary method of sampling used to collect data is through convenience sampling. This method has been shown to provide samples with a significantly diverse demographic makeup. The online parent report survey is the most used method for understanding children's language vocabulary development due to its low cost and wide reach, exceeding the sample size typically attained through naturalistic observation or experimental tests. Parent reports also utilize the extensive knowledge of parents or primary caregivers about their child's behaviour, enabling more comprehensive research than what a researcher alone can obtain. These reports are useful in providing assessments of children's development during the critical age range when language begins to emerge until around 30 months of age. The vocabulary lists used in assessing their development present a variation in development across various ages and abilities based on words that the child can comprehend or comprehend and produce, as determined by the parent. The bank of words includes both semantic and syntactic categories, providing a comprehensive list of terms that most children will know.

The analysis of this paper makes use of the R programming language (R Core Team 2020) for statistical computations and visualizing data. The tidyverse package (Wickham et al. 2019) is installed to gain access to other important R packages, including the dplyr package (Wickham et al. 2023) used to manipulate and clean data, the readr package (Wickham, Hester, and Bryan 2023) to read and import data, the here package (Müller 2020) to create a path to specific saved files, the ggplot2 package (Wickham 2016) to create the data visualizations, and the modelsummary package (Arel-Bundock 2022) to create summary tables.

2.2 Variables of Interest

Our analysis is specifically interested in investigating the production of vocabulary words in American Sign Language. In this paper, we use a dataset focused solely on ASL that is comprised of 3366 observations, where each row represents an evaluation of a child and each column indicates a specific variable, which could be a demographic factor or a response from the parent regarding a particular item (Frank et al. 2016a). Moreover, this paper focuses on the analysis of several variables, including the unique identification number of the child, their age in months, the specific vocabulary item they are being assessed on, the item's category, and the overall assessment of their ability to understand and/or produce the vocabulary item. It is important to note that the 3366 available observations from the data on learning ASL were collected for only six unique children.

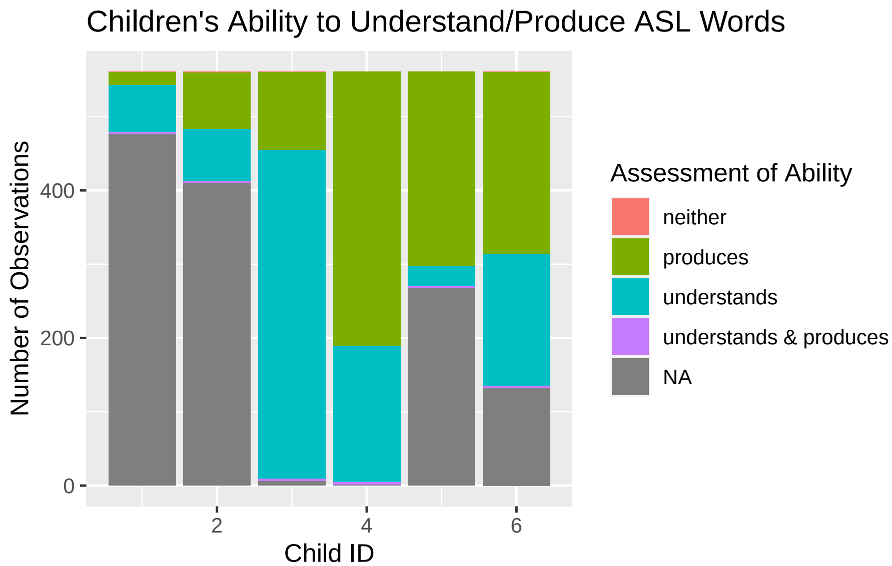

Figure 1: Each child's overall assessment by American Sign Language (ASL) word, with their identification number arranged in ascending order of age.

The vocabulary development of children in American Sign Language is presented in Figure 1, providing an overview of their overall assessment. Possible assessments for the child's language abilities include understanding the item, being able to understand and produce the item, the inability to understand and produce the item, or the assessment report not being applicable. Notably, according to the database, children who can produce a vocabulary item are also inferred to understand it as well. Additionally, this graph provides a visual representation of each child's abilities, with their identification number arranged in ascending order according to age.

At 12 months old, the youngest child (Child 1) primarily comprehends a few ASL words and produces only a limited number. In contrast, the next child in age (18 months old) is capable of producing a greater number of items. A similar trend is observed for Child 3 (21 months old) as they are able to understand a greater amount of ASL vocabulary items and produce a slightly higher number than the previous child. In comparison to Child 3, the data on Child 4 (28 months old) indicates a significant increase in production capabilities. As we move to Child 5 (32 months old) and Child 6 (33 months old), they exhibit the ability to produce more items than the items that they are only able to comprehend. However, it is important to note that these two children have a higher number of not applicable (NA) responses in comparison to Child 3 and Child 4. The limitations of these NA responses are discussed in Section 5.4.1 of this paper.

Table 1: Summary statistics table of the children's ASL vocabulary assessment.

		N	%
Production Ability	NO	4	0.1
	YES	2070	61.5
	NA	1292	38.4
Comprehension Ability	NO	1085	32.2
	YES	989	29.4
	NA	1292	38.4

The summary statistics of the total American Sign Language vocabulary items produced and comprehended across the 3366 observations by the six unique children learning ASL are presented in Table 1. In this table, the production ability category comprises items that children were capable of both comprehending and producing. Conversely, the comprehension ability category includes items that the children were only able to understand and not produce. Within this table, the "N" column represents the total number of observations for each variable, while the "%" column displays the percentage proportion of children for each given variable. Notably, more than 50 percent of the observations show a vocabulary item that was both understood and produced by a child. Additionally, about 30 percent of the observations indicate the number of ASL vocabulary items that children were able to comprehend but did not produce. Moreover, approximately 40 percent of the observations had NA responses for the children's vocabulary development assessment.

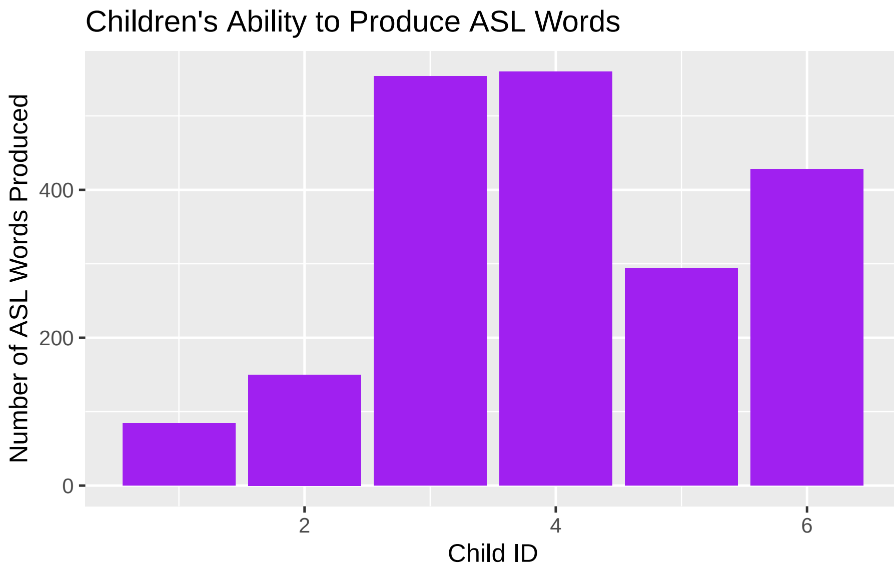

Figure 2: Each child's amount of produced ASL words, with their identification number arranged in ascending order of age.

Figure 2 focuses on production ability, displaying each child's total number of American Sign Language vocabulary items that they are able to comprehend and produce. Notably, the identification number for each child is organized in ascending order based on their age. From this graph, we can observe a negative skewness pattern where the graph is skewed to the left, indicating a gradual increase at the start of the graph and a significant increase as the graph moves toward the center. The data visualization demonstrates that the age of the child that had the least amount of vocabulary items produced is 12 months old (Child 1). Contrastingly, Child 3 (21 months old) and Child 4 (28 months old) have the highest number of ASL vocabulary items produced. Near the end of the graph, where the children with higher ages (32 months and 33 months old) lie, there is a slight decline in the number of produced ASL vocabulary items in comparison to the two former children.

3 Model

Through the data analysis of this paper, we discovered that there is a correlation between a child's age and the number of American Sign Language vocabulary items that they are able to produce, which has implications for their overall vocabulary development in ASL. To further analyze and infer the age timeline at which children can produce a certain number of vocabulary words in sign language, we constructed a linear regression model.

Thus, we estimate the following model:

$$Y_i = \beta_0 + \beta_1 Age_i \tag{1}$$

In Model 1:

- Y_i is the change in the number of produced ASL vocabulary items for each additional month of age, given every one unit change in age.
- β_0 is the coefficient for intercept.
- β_1 is the coefficient for the rate at which the number of produced ASL vocabulary items changes with respect to age in months.

The independent variable of the model is the age of the child, measured in months. The dependent variable is the number of ASL vocabulary items that the child is able to produce.

3.1 Model Justification

We anticipate that the development of vocabulary word production, whether it be for a spoken language or a visual language, would be close to linear. This is because we assume that as a child gets older in age, their cognitive abilities improve, leading to an increased ability to produce more vocabulary words. Before building the model, we estimate the relationship between the number of produced vocabulary items (dependent variable) and age in months (independent variable) using the known values of these variables. These values were obtained from Figure 2, specifically for Child 2 and Child 3, whose ages are 18 months and 21 months respectively. The reason for calculating the relationship between the number of produced vocabulary items and age in months based on Child 2 and Child 3 is that they are the closest in age to each other. We chose to exclude Child 5 and Child 6, who are 32 months and 33 months old, respectively, when considering which known values to calculate because they are close to the age boundaries at which children are able to engage in face-to-face activities, typically around 30 months old. Therefore, it is more reasonable to estimate the relationship based on the rate of the change between Child 2 and Child 3, as they fall within the crucial age range when language emergence is common. Additionally, since the available data on vocabulary development in ASL only involved six unique children and their ages, we simulated additional data to fill in the gaps. Since we needed data that spanned the spectrum of ages between 12 months to 36 months old, we used the simulated data and the estimated relationship previously mentioned to generate inferential statistics by using the model.

4 Results

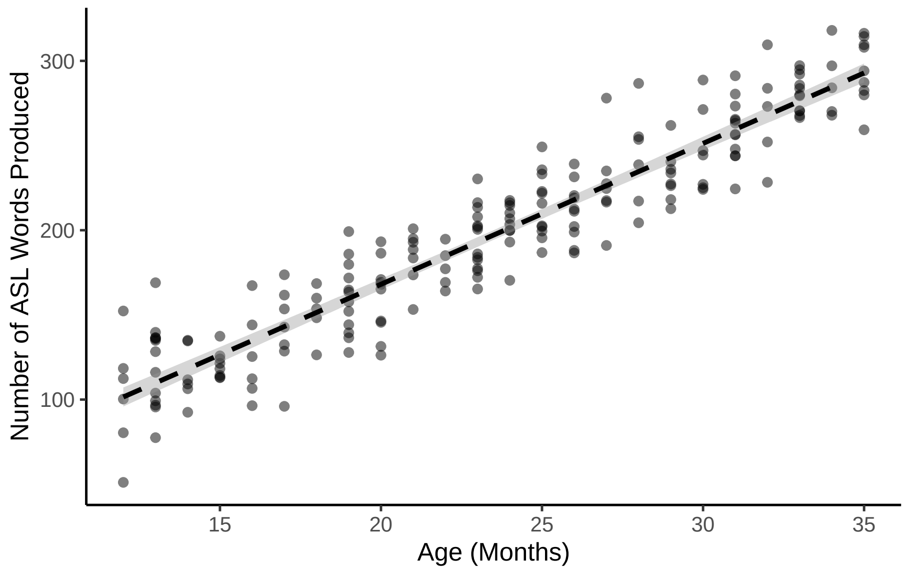

Figure 3: Linear regression model of the number of produced ASL words by the range of children's ages during their early developmental years.

Model 1, as illustrated in Figure 3, displays the correlation between a child's age in months and the number of ASL words they can produce at a given age. The scatterplot depicts the general trend of vocabulary growth as a child grows in age, further emphasized by the regression line. The model indicates that as the age of a child increases, the number of vocabulary items that that child produces increases as well. The regression line allows for the prediction of future outcomes based on the known data available from the dataset. By utilizing this model, we can estimate when a child will be able to produce a certain number of ASL vocabulary words or phrases at a certain age.

Table 2: Model summary of the number of produced ASL words by the range of children's ages during their early developmental years.

	(1)
(Intercept)	1.695
	(5.208)
age_in_months	8.319
	(0.213)
Num.Obs.	200
R2	0.886
R2 Adj.	0.885
AIC	1776.6
BIC	1786.5
Log.Lik.	−885.283
RMSE	20.24

Table 2 shows the coefficients of the predictor variables of Model 1. We are concerned with the top half of this table, as it provides the values of the coefficient representing the intercept as well as the coefficient representing the rate at which the number of produced ASL vocabulary items changes with respect to age in months. The standard errors of the estimated coefficients are also included in brackets. This is also provided visually in Figure 3 by the means of the grey area surrounding the regression line, presenting the degree of error in the prediction. As seen in Figure 3, there is no intercept in the model. Under these circumstances, the intercept can be disregarded, as there is no value of 0 within the range of the independent variable, age in months. The spectrum of months explored in this model is between the early age of 12 months old to 36 months old. It is important to note that we specify the estimated relationship between the independent variable and the dependent variable. We, therefore, took thee estimated relationship to the available data in order to estimate the coefficients of the model. Hence why the coefficient value for *age_in_months* is close to our estimated relationship value.

5 Discussion

5.1 Learning sign language, especially at an early age, creates a more accessible world for the deaf and hard-of-hearing community

Learning sign language at an early age expands children's understanding of diverse individuals and communities, fostering their ability to be more empathetic and accommodating towards accessibility. Being able to communicate through sign languages, such as American Sign Language, provides more accessible and effective opportunities for individuals who are deaf or hard of hearing ("What Is American Sign Language?" 2021). Sign language is a highly accessible form of communication that enables individuals with hearing impairments to engage in conversations on equal grounds with others, ultimately accommodating diversity and universal inclusion.

5.2 At early ages, children have strong capabilities of learning multiple languages, including sign language

The timeline of language development for spoken languages during the emerging years of children has a similar growth trend to the data analysis in this paper, with a more significant upward trend. While children learning to speak a spoken language are able to produce their first words between the ages of 6 to 11 months old ("Age-Appropriate Speech and Language Milestones," n.d.), those learning to communicate through sign language produce one-word vocabulary by the age of 12 months ("Language Development Milestones in Sign Langauge from Age 0 to 5," n.d.). Afterwards, they both follow the same trend of producing 2-word phrases around the ages of 18 to 24 months old and by 2-3 years of age, they are able to use 3-word phrases. These highlighted milestones indicate that the vocabulary development of spoken languages shows a similar linear growth trend to the vocabulary development of sign languages.

Early learning of sign language is beneficial as it helps children have a better comprehension of the language while simultaneously improving their cognitive and social development. Research suggests that the early years of a child's life are the most crucial in terms of the development of their language skills ("What Is American Sign Language?" 2021). Children at early ages have the capability to learn multiple languages at a time without confusion or any sort of delay in their development, as research has shown that their brains are able to handle dual language development ("American Sign Language/English Bilingual and Early Childhood Education," n.d.). They are able to expand their communication and linguistic skills as well as provide access to a language that meets the needs of deaf individuals.

5.3 Learning sign language enhances visual skills and improves cognitive abilities

Expanding on the cognitive and social benefits that coincide with the development of sign languages in children, it also enables children to enhance their visual skills and cognitive abilities. Sign languages allow children to communicate and express themselves effectively. For instance, when a child is unsure of a sign in the language, they are able to work around this barrier by using the letters of the alphabet to express what they are trying to say ("The Benefits of Using Sign Language in Early Child Development," n.d.). Additionally, because sign language involves the physical movement of the body, the use of muscle memory helps increase a child's vocabulary ("The Benefits of Using Sign

Language in Early Child Development," n.d.). It helps perform words automatically without conscious effort or attention, demonstrating a retained memory of vocabulary. Alongside movement, sign languages are also visual languages. This helps in enhancing the visual learning experience of children during their developmental stages ("The Benefits of Using Sign Language in Early Child Development," n.d.). With this, they are able to see words in motion and retain both the motion of the word or phrase and the vocabulary word itself. Teaching children sign language, whether they are experiencing hearing loss or are able to communicate via speech, can have a positive impact on their language development as well as their social interactions and relationships.

5.4 Weaknesses & Next Steps

5.4.1 Weaknesses

The limitations of using data from Wordbank are that its primary focus is on typical development and its datasets focus on monolingual acquisition. In addition, the sample of languages is restricted by data accessibility. Additionally, there are a lot of NA responses within the dataset. This invites the problem of missing data, which can reduce the sample size and have an overall reduced statistical power. Missing data could also introduce bias into the analysis, especially with the main data collection being the survey reports from parents or primary caregivers. This can affect the representativeness of the sample. Moreover, due to the limited number of unique children for the data on American Sign Language development, a large number of observations had to be simulated for the model and its analysis.

5.4.2 Next Steps

A possible next step could be to explore how different categories of vocabulary words are developed over the age range and their varying levels of comprehension and production of those words. With that being said, we could also extend our linear regression model to be a multiple linear regression model, to accommodate the different categories. Regarding the NA responses of the dataset, what could be done to improve the overall analysis would be to fill in missing values with estimated values based on other the responses. This would be beneficial for both analyzing trends of the data and generating linear regression models. Additionally, since we based our timeline of milestones for learning spoken languages on secondary research, providing data visualizations and a model to compare and contrast the development of visual languages with spoken languages would enhance the analysis of this paper.

References

"Age-Appropriate Speech and Language Milestones." n.d. https://www.stanfordchildrens.org/en/topic/default?id=age-appropriate-speech-and-language-milestones-90-P02170; Stanford Medicine: Children's Health.

"American Sign Language/English Bilingual and Early Childhood Education." n.d. https://www.colorincolorado.org/article/american-sign-languageenglish-bilingual-and-early-childhood-education; Colorin Colorado.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Frank, Michael C., Matthew Braginsky, Daniel Yurovsky, and Virginia A. Marchman. 2016a. "Wordbank Instrument Data - Language: American Sign Language, Form: FormA." http://wordbank.stanford.edu/data?name=instrument_data.

———. 2016b. "Wordbank: An Open Repository for Developmental Vocabulary Data." http://wordbank.stanford.edu/data.

"Language Development Milestones in Sign Langauge from Age 0 to 5." n.d. https://www.handspeak.com/learn/416/; Handspeak.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"The Benefits of Using Sign Language in Early Child Development." n.d. https://learningjournals.co.uk/benefits-of-sign-language-in-child-development/; Learning Journals.

"What Is American Sign Language?" 2021. https://www.nidcd.nih.gov/health/american-sign-language#:~:text=American%20Sign%20Language%20(ASL)%20is,of%20the%20hands%20and%20face; National Institute on Deafness; Other Communication Disorders.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.