Computational Linguistics and Chinese Language Processing
Vol. 25, No. 1, June 2020, pp. 81-102

Linguistic Input and Child Vocalization of 7 Children from 5 to 30 Months: A Longitudinal Study with LENA Automatic Analysis

Chia-Cheng Lee*, Li-mei Chen+, and D. Kimbrough Oller#

Abstract

This study examined longitudinal changes in linguistic input, conversational turns, and child vocalizations in Chinese-speaking families using the computerized LENA (Language Environment Analysis) software, a system that captures audio data in children's natural environment and parses out speech data automatically. All-day home recordings (11-16 hours) from seven typically developing Chinese-learning children (two males and five females) at the ages of 5, 10, 14, 21, and 30 months were analyzed. Adult word count (AWC), conversational turn count (CT), and child vocalization count (CV) of 70 recordings (i.e., 7 children x 5 ages x 2 recordings) were retrieved from the LENA software. These recordings included times when families were asleep. As a result, the present study also compared the results with and without LENA-determined silence time (i.e., quiet and sleep time). The results showed that the percentage of silence in the recordings decreased with age, indicating that the children's awake time increased as they age. When the children were awake, they listened to an average of 1734 adult words, engaged in 39 conversational turns, and produced 150 vocalizations per hour from 5 to 30 months of age. The CV and CT increased with age, while the AWC did not show a clear pattern, which was similar to English normative estimates from Gilkerson and Richards (2008). The CT was also found to be a more effective contributor to the number of CV than AWC, indicating that speech produced in temporal proximity

* Department of Speech and Hearing Sciences, Portland State University, USA
E-mail: jiajiajanejane@gmail.com
+ Department of Foreign Languages and Literature, National Cheng Kung University
E-mail: leemay@mail.ncku.edu.tw
The author for correspondence is Li-mei Chen.
School of Communication Sciences and Disorders, The University of Memphis, USA
E-mail: koller@memphis.edu

to children's vocalizations or directed to children played an important role in eliciting child vocalizations.

Keywords: LENA, Adult Word, Conversational Turn, Child Vocalization, Longitudinal Study, Cross-language Comparison

1. Introduction

Child speech samples have traditionally been collected by visiting children's homes or inviting families into a research laboratory. LENA (Language Environment Analysis) software, a system that collects audio data without research assistants' presence and parses out audio data into several categories automatically, was developed in 2004 in the United States (LENA Research Foundation, 2020). The software has been used for observing English-speaking individuals (Gilkerson & Richards, 2008; Greenwood, Thiemann-Bourque, Walker, Buzhardt & Gilkerson, 2011; Suskind *et al*., 2013), Chinese-speaking families (Gilkerson *et al*., 2015; Lee, Jhang, Relyea, Chen & Oller, 2018; Zhang *et al*., 2015), preterm infants (Caskey, Stephens, Tucker & Vohr, 2011, 2014), multilingual speakers (Liu & Kager, 2017; Oller, 2010; Orena, Polka & Srouji, 2018), individuals with disorders (Ambrose, VanDam & Moeller, 2014; Charron *et al*., 2016; Oller *et al*., 2010; Thiemann-Bourque, Warren, Brady, Gilkerson & Richards, 2014; VanDam, Ambrose & Moeller, 2012; Warren *et al*., 2010), and older adults (Li, Vikani, Harris & Lin, 2014). The number of studies on the quantity of linguistic input, conversational turns, and child vocalizations in Chinese-speaking home environments have been limited. The present study observed changes in the quantity of linguistic input, conversational turns, and child vocalizations which occur between 5 and 30 months of age in Chinese-speaking families using LENA.

Research has shown that linguistic input, including the quantity and quality of caregiver speech and turn taking sequences, plays an important role in the child's vocal development (Caskey *et al*., 2011; Hart & Risley, 1995; Rowe, 2012; Suskind *et al*., 2013). This in turn serves as a strong predictor of their later vocabulary growth (Hart & Risley, 1995; Ramírez-Esparza, García-Sierra & Kuhl, 2014). Studies have also found that early vocal production is associated with future speech and language development. Rescorla *et al*. (2000) indicated that some children who were identified as late talkers at two years of age continued to exhibit language delay and were identified as children with Specific Language Impairment at three years of age. Gilkerson *et al*. (2018) also showed that school-age language and cognitive outcomes (9-13 years old) and quantity of adult talk and adult-child interaction during 18 to 24 months of age are related.

1.1 Linguistic Input and Conversational Turn

Linguistic input from adults or siblings is identified as one of the largest influences on children's verbal performances, including that of preterm infants (Caskey *et al*., 2011). Children understand five times more words than the words they produce (Ingram, 1989), suggesting that a substantial number of words need to be heard before a child speaks. Roy *et al*. (2009) reported that adult word input frequencies and age of acquisition of words is highly correlated. Adult word input between 10 and 36 months of age has been found to be related to a child's IQ at 3 years (Hart & Risley, 1995). Gilkerson and Richards (2009) also found that children who scored higher on language assessments tended to have talkative parents. The number of words parents spoke to children between two and six months of age predicted language ability at two years of age. Parents who earned at least a bachelor's degree talked more to their children than less educated parents. Also, first-born children were spoken to more than later born children.

Children may be at risk of learning languages if they do not have sufficient language exposure (Velleman & Vihman, 2002). Many scholars have claimed that language acquisition takes place even when the linguistic input that children are exposed to is addressed to them indirectly (Akhtar, Jipson & Callanan, 2001; Oshima-Takane, 1988; Oshima-Takane, Goodz & Derevensky, 1996). Other scholars argued that speech addressed directly to children has a stronger effect on children's language learning (Oller, 2010; Pearson, Fernandez, Lewedeg & Oller, 1997; Shneidman, Arroyo, Levine & Goldin-Meadow, 2013; Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013). The same phenomenon has been posited by Shneidman *et al*. (2013) and Shneidman and Goldin-Meadow (2012), who found that direct speech has a more important role in early word learning than indirect speech in children who grew up in communities where indirect speech was the major linguistic input.

In addition to receiving speech and language input, children also respond to the input (Hart & Risley, 1995). Mother-child vocal interactions have been discussed in several studies (Gratier *et al*., 2015; Gros-Louis, West, Goldstein & King, 2006; Jaffe *et al*., 2001). From 3 to 4 months of age, infants start to use pragmatic, semantic, and syntactic factors to predict when a conversational turn will end and begin (Gratier *et al*., 2015). However, studies on linguistic input and conversational turn-taking in Chinese-speaking environments, especially vocalizations produced in home environments, are, as of yet, few in number. Studies investigating the relationship among linguistic input, conversational turns, and children's vocalizations should shed some light on our understanding of the relationship between different types of linguistic input and language development.

1.2 Assessing Vocal Development Using an Automated Approach

Although the LENA system was mostly utilized in American-English environments, the system has yielded valid and reliable speech and language estimates in other languages (French: Canault, Le Normand, Foudil, Loundon & Thai-Van, 2016; Spanish: Weisleder & Fernald, 2013, Chinese (Mandarin and Shanghai dialect): Gilkerson *et al*., 2015; Zhang *et al*., 2015; Korean: Pae *et al*., 2016; Dutch: Busch, Sangen, Vanpoucke & van Wieringen, 2018; Vietnamese: Ganek & Eriks-Brophy, 2018). After comparing Chinese speech samples analyzed by the LENA system with the same samples transcribed by a native Chinese transcriber, Gilkerson *et al*. (2015) indicated that the validity of the LENA system in identifying and estimating adult words, child vocalizations, and conversational turns is reasonably accurate. Zhang *et al*. (2015) observed 22 Chinese-speaking families and their typically developing children between 3 and 23 months of age in Shanghai for a period of 6 months. A total of 19 recordings were made by each family. The 22 families were divided into two groups based on the speech output of the first three recordings. One group of families had fewer adult words (Group A), while the other group had a higher rate of adult words (Group B) in their first three recordings. The authors provided monthly feedback to the families regarding strategies to increase their linguistic input to and interaction with their children. The results overall showed that adult words and conversational turns increased during the first three months, but decreased during the last three months. However, Group A showed increased number of adult words in the last few recordings, which was not observed in Group B. The study indicates that the LENA system can be used to track children's vocal, speech, and language development and/or treatment progress. The authors also found that their number of conversational turns correlated positively with the MacArthur-Bates Communicative Development Inventories - Verbal (Fensen *et al*., 2007) and Minnesota Child Developmental Inventory Expressive Language (Ireton, 1992) scores for the change from baseline to 3 months. LENA estimates have also shown reliable and valid results when compared with scores of standardized assessments (Richards *et al*., 2017), including – Preschool Language Scale – 4th Edition (Zimmerman, Steiner & Pond, 2002) and the Receptive-Expressive Emergent Language Test – 3rd Edition (Bzoch, League & Brown, 2003).

Table 1 shows adult word count (AWC), conversational turn count (CT), and child vocalization count (CV) per hour from various ages, settings, and population. Depending on the children's age and the recording environment, children received different linguistic input and produced different number of words. AWC ranged from 889 to 1966. CT ranged from 17 to 75. CV ranged from 73 to 188 per hour. Gilkerson and Richards (2008) examined a corpus of spontaneous speech data in English-speaking families and created normative estimates for CV and CT each month when children were between 2 and 48 months of age. Here only values measured at 5, 10,14, 21, and 30 months are listed in Table 1.

Table 1. Studies reported AWC, CT, and CV per hour in families with 0-3-year-old children

First author & Year	Population, *n* (male/female)	Age	Language	AWC per hour	CT per hour	CV per hour
Ambrose (2014)	Hard of hearing, *n*=28 (10/18)	12-36 months (mo)	English	1429	59	Not Applicable (NA)
Gilkerson (2008)	Typically Developing (TD), *n*=329 (167/162)	2-48 mo	English	NA	5 mo 17 10 mo 23 14 mo 27 21 mo 36 30 mo 40	5 mo 73 10 mo 95 14 mo 102 21 mo 145 30 mo 184
Greenwood (2013)	TD, *n*=30 (NA/NA)	12-20 mo	English	1095	38	143
Thiemann-Bourque (2014)	Down syndrome (DS), *n*=9 (3/6), and age- and gender-matched TD, *n*=9 (3/6)	9-54 mo, young DS 9-11 mo, old DS 25-54 mo	English	Yong DS 889 Old DS 1044 TD NA	Young DS 18 Old DS 19 TD 44	Young DS 102 Old DS 64 TD 179
Warren (2010)	Autism, *n*=26 (22/4), and age- and gender-matched TD *n*=78 (66/12)	16-48 mo	English	Autism 1079 TD 1138	Autism 35 TD 4	Autism 134 TD 188
Zhang (2015)	TD, *n*=22 (10/12)	3-23 mo	Shanghai dialect and Mandarin	Baseline 1758 1 mo 2174 1-3 mo 1966 4-6 mo 1711	Baseline 63 1 mo 75 1-3 mo 66 4-6 mo 56	NA

1.3 The Present Study

Because of the laborious coding required for estimating linguistic input from the ambient environment, studies focusing on child speech development are usually based on a limited set of recordings. To our knowledge, only three studies (Gilkerson *et al*., 2015; Lee *et al*., 2018;

Zhang *et al*., 2015) reported observations in Chinese-learning children's natural environments using LENA. In view of this, LENA was adopted for data collection and processing in the present study. This paper explores the relationship among children's vocalization, the linguistic input children received, and amount of interaction adults and children had per hour (e.g., total number of AWC/total length of a recording). However, the recordings included times when families were asleep. Thus, the present study investigated the research questions using the total length of the recording without LENA-determined silence time (i.e., quiet, sleep time) to calculate another set of average numbers of AWC, CT, and CV per hour (e.g., total number of AWC/(total length of a recording without silence time in the recording)). Periods of silence were removed to ensure that the analysis only included times when children were most likely to be awake. Analyzing results by removing periods of silence time from LENA recordings has also been reported in several other studies (Marchman, Martínez, Hurtado, Grüter & Fernald, 2017; Sacks *et al*., 2013). Since children at 0-2 years old sleep an average of 12.7 hours a day and children at 2-3 years old sleep an average of 12 hours a day (Galland, Taylor, Elder, & Herbison, 2012), the results of the present study could have been influenced by long sleeping times. Therefore, the present study aimed to compare the results when silence time was included with the results when silence time was removed from the analyses.

The present study investigated the following questions:

1. Do adult word count (AWC), conversational turn count (CT), and child vocalization count (CV) increase as children grow older?
2. Are there different patterns in AWC, CT, and CV when LENA-determined silence time is removed?
3. Are both AWC and CT effective contributors to the number of CV at 5, 10, 14, 21, and 30 months?
4. Do AWC, CT, and CV show cross-language differences?

2. Methods

2.1 Participants

Seven Chinese-speaking families and their children (two males and five females) participated in the study. The families lived in Tainan, Taiwan, an environment where Mandarin Chinese and Southern Min (Taiwanese) were mostly spoken. All the children were born full-term without hearing or neurodevelopmental disorders. Table 2 shows demographic information of the participants.

Table 2. Demographic information of the participants

Child	Gender	Birth order	Mother's education
A	F	1st	M.A.
B	F	1st	B.A.
C	F	1st	B.A.
D	F	1st	B.A.
E	M	1st	B.A.
F	M	2nd	B.A.
G	F	2nd	B.A.

2.2 Recording Procedure

The digital language processor (DLP), a recording device developed along with the LENA Pro system (LENA Research Foundation, 2020), was used to collect data. Before each recording session started, a child wore a specially designed vest with a DLP (Figure 1). The caregiver turned the DLP on to start a recording session and switched the DLP off after 16 hours of recording. The recording file was automatically uploaded and processed (Figure 2) once the DLP was connected to a computer with the LENA Pro software. The LENA Pro software identified speech and other sounds from each recording and generated counts at 5-minute, hour, day, and month intervals. The authors retrieved the counts/reports (Figure 3) from the software for further analysis.

Figure 1. The LENA digital language processor (DLP) placed in the pocket of a vest

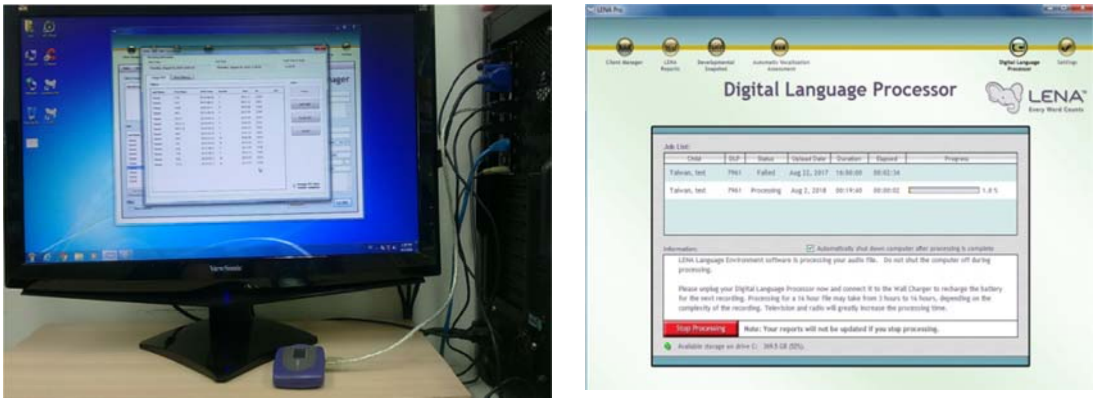

Figure 2. Data transfer from a DLP to the LENA Pro software

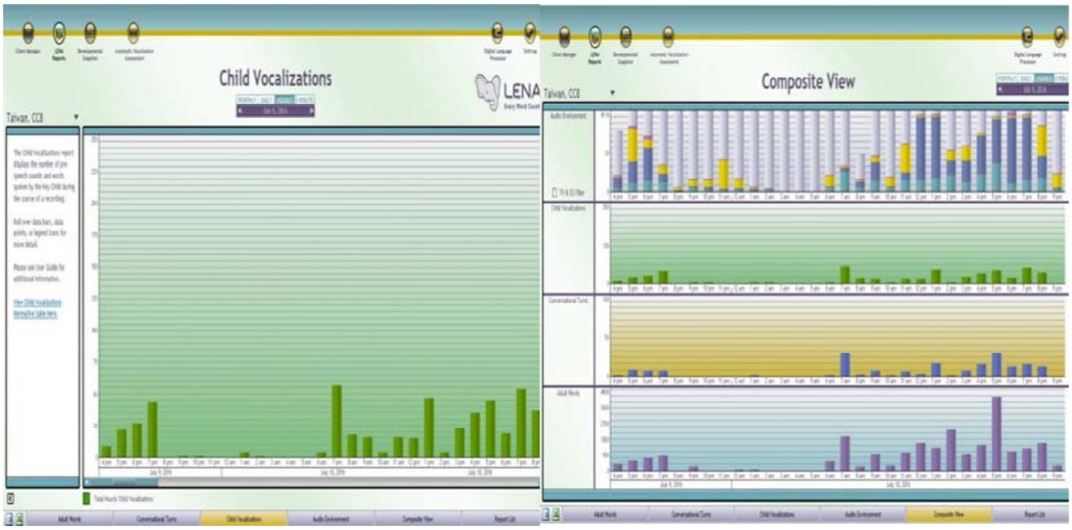

Figure 3. Reports from the LENA Pro software

A set of two recordings were made at each age: 5, 10, 14, 21, and 30 months old. A total of 70 recordings were analyzed (i.e., 7 children x 5 ages x 2 recordings). All the recordings were 16 hours in length except for 6 of the recordings due to insufficient power of the device used on the recording day. The 6 recordings were between 11 and 14 hours in length.

2.3 Data Processing by the LENA Software

The audio data was processed and categorized by the LENA Pro software into eight sound categories: (1) the key child who wore a vest with the DLP, (2) other child, (3) adult male, (4) adult female, (5) overlapping sounds, (6) noise, (7) electronic sounds (e.g., TV), and (8) silence (i.e., silence, quiet, or vegetative sounds such as sneezes, coughs, or snores). Each category was further identified as clear and unclear (i.e., quiet and distant) subcategories. After the eight sound categories were identified, the LENA system determined adult word count (AWC), communication turn count (CT), and child vocalization count (CV).

2.3.1 Adult Word Count (AWC)

AWC measured the total number of words spoken around the key child. Using acoustic features in speech signal (e.g., formants, pitch, segment duration, silence duration), adult sounds were identified as phones using American-English phone parsing models. Speech segments were identified based on differential acoustic energy patterns, and no specific adult words were identified. AWC included both speech directed to the key child and speech directed to others. In Mandarin Chinese, one syllable represents one spoken syllable, whereas one word may contain one or more spoken syllables. For example, 窗戶 chuang hu (*window*) has two spoken syllables but counts as one word. Gilkerson *et al*. (2015) compared syllable count (e.g., 窗戶 chuang hu = two syllables) and word count (窗戶 chuang hu = one word) transcribed by a trained native Chinese human transcriber with AWC and found that both comparisons showed valid and reliable estimates of adult word count. The authors suggested that since the comparisons were both reliable, researchers can use LENA-determined AWC (syllable count) in future studies. The authors also indicated that since all languages have phonemes and syllables, and the acoustic features of consonants and vowels are similar across languages, using acoustic information to estimate adult word count should not be affected by language differences.

2.3.2 Conversational Turn Count (CT)

Conversational turn count (CT) refers to the total number of conversational turns the child engaged in with other speakers. A conversational turn is defined as a child speaking and an adult or a child responding, or an adult or a child speaking and the child responding within 5 seconds. Both intentional and unintentional vocal production and responses can be counted as turns.

2.3.3 Child Vocalization Count (CV)

Child vocalization count (CV) is the total number of speech-related vocalizations the child produces. A CV would be identified if there was a 300 millisecond or longer vocal break between the key child's vocalization. Cries, laughs, and vegetative sounds such as sneezes, coughs were excluded from child vocalization count. Similar to AWC, the LENA system did not identify specific words or syllables in utterances. If a child says "ma" or "I want that I want that I want that" without pauses between words, each utterance is counted as one vocalization.

2.4 Data Analyses

Five categories retrieved from the LENA reports were used for further analyses in the present study: (1) the length of each recording, (2) adult word count (AWC), (3) communication turn count (CT), (4) child vocalization count (CV), and (5) length of silence in each recording. The total number of words or sounds in each recording may differ depending on the length of the recording. Since the length of each recording was different, the average number of AWC, CT, and CV per hour retrieved from each recording was first calculated. Next, the average number of AWC, CT, and CV per hour retrieved from each recording without silence were calculated. Two sets of statistical measures were then analyzed. First, six one-way repeated measure ANOVAs were performed to explore whether there were any changes in the three variables (the average number of AWC, CT, and CV per hour) across time as well as when silence was included or excluded. Next, ten multiple regressions were performed at the ages of 5, 10, 14, 21, and 30 months to examine how much AWC and CT contribute to CV at each age and whether or not silence was included.

3. Results and Discussion

3.1 Changes of AWC, CT, and CV Overtime

Figure 4A shows the average number of adult word count (AWC), conversational turn (CT), and child vocalization (CV) per hour and their standard deviations from the recordings made at 5, 10, 14, 21, and 30 months. The average number of AWC per hour shows an increase from 5 to 10 months and a gradual decrease from 10 to 30 months. However, the differences among the five ages are not statistically significant, which is similar to the finding of Gilkerson and Richards (2008). The authors stated that AWC and chronological age in English-speaking families were not significantly correlated. The results in the present study also showed that the number of child vocalizations increased slowly with age, even when the child received a fair amount of linguistic input from the environment. That is, children heard an average of 412 to 752 adult words per hour from 5 months to 30 months old. However, the average number of child vocalizations only increased from 27 to 90 vocalizations per hour from 5 to 30 months.

The average number of CT per hour also shows a gradual increase from 5 (5 per hour) to 30 (23 per hour) months. The differences among the five ages are statistically significant [$F(4, 24) = 3.318$, $p < .05$]. A post hoc analysis indicates that the average number of CT per hour at 21 months (18 per hour) is significantly higher than at 5 months (5 per hour) [$t(6) = 3.716$, $p < .05$]. The increased number of CT indicates that the adults became more and more responsive to their children’s utterances, and vice versa. The adults may have initiated the conversation when they thought that their children were ready to talk, or responded to their child utterances right away. The children may have also learned to gain other people's attention by producing sounds. Or, they may have learned to respond to adults' speech right away as they grew older.

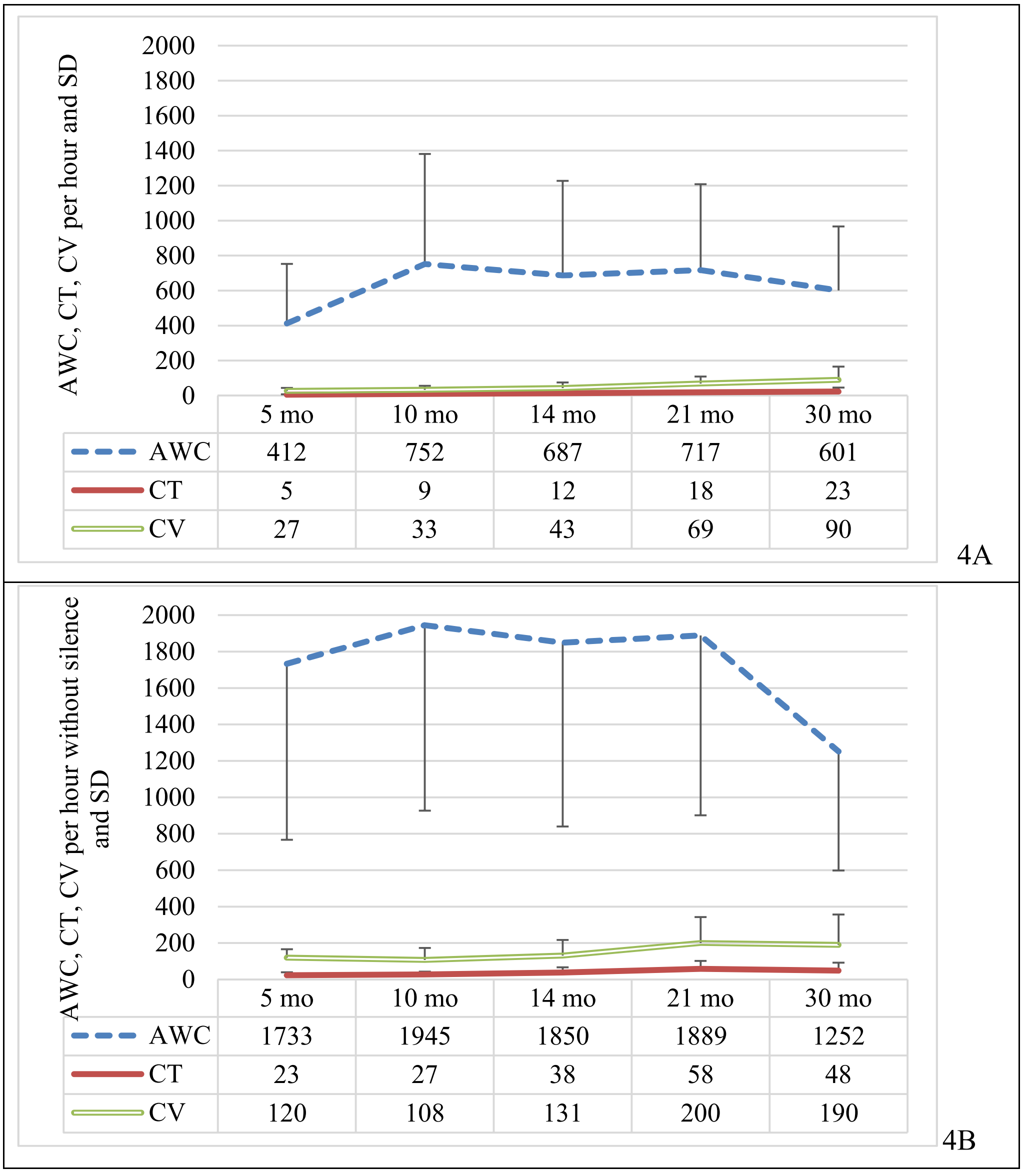

Figure 4. Average adult word (AWC), conversational turn (CT), child vocalization (CV) per hour with and without silence and standard deviations at 5, 10, 14, 21, and 30 months

3.2 Changes of AWC, CT, and CV Overtime after Removing Silence

Periods of silence were removed from recordings to ensure that only the times when children were most likely to be awake were included in the analysis. Figure 4B shows the average number of AWC, CT, and CV per hour and their standard deviations after removing the periods of LENA-determined silence from the recordings. The standard deviations of the average AWC per hour was high at all five ages as shown in both Figures 4A and 4B. However, the variability across families is even higher after the periods of silence were removed. The percentage of silence (i.e., (silence time/total length of recording) x 100) decreased with age (i.e., 5 mo: 73%, 10 mo: 66%, 14 mo: 62%, 21 mo: 59%, 30 mo: 48%), which was in line with Galland's *et al.* (2012) finding that children's sleep time decreased with age.

As expected, the mean number of the three variables was at least twice as high without silence as with silence. Without silence time, the average number of CT and CV per hour also gradually increased from 5 (CT: 23; CV: 120 per hour) to 30 months (CT: 48; CV: 190 per hour). But, the differences among the five ages were not statistically significant. The average number of AWC per hour showed an increase from 5 (1733 per hour) to 10 (1945 per hour) months and a gradual decrease from 10 to 30 (1252 per hour) months. Yet, the differences among the five ages were not statistically significant either. Also, the average number of CT per hour was significantly different across ages before silence was removed, but was not significant after silence was removed. The average number of CT (i.e., increased with age), and the periods of silence (i.e., decreased with age) may account for the change.

In addition, the AWC and CT in the present study from the data across the five ages with silence removed (AWC: 1734; CT: 39 per hour) were more similar to Chinese-speaking data from Zhang *et al*. (2015) (AWC baseline: 1758; CT baseline: 63 per hour) than the results with silence included (AWC: 634; CT: 14 per hour). Zhang *et al.*'s (2015) results were more similar to results when silence was excluded in the present study because the authors instructed their Chinese-speaking families to record for 12 hours during the daytime. The finding also suggests that LENA-determined silence was identified as reasonably accurate.

3.3 Relationships among AWC, CT, and CV

Multiple regressions were performed at each age to explore the relationship among AWC, CT, and CV. The results showed that the numbers of AWC and CT could predict the numbers of CV at 10 months and 30 months. At 10 months, the results of the regression indicated that the model explained 88.1% of the variance and that the model was a significant predictor of the number of CV, $F(2,4) = 23.306$, $p = .006$. While the number of CT contributed significantly to the model ($B = 3.677$, $p = .003$), the number of AWC did not ($B = -.008$, $p = .222$). That is, the increase of one unit of CT could contribute to the increase of 3.677 units of CV. At 30 months, the results of the regression indicated that the model explained 95% of the variance and that the model was a significant predictor of the number of CV, $F(2,4) = 57.9$, $p = .001$. While the number of CT contributed significantly to the model (B = 3.899, $p = .002$), the number of AWC did not (B = -.044, $p = .266$). That is, the increase of one unit of CT could contribute to the increase of 3.899 units of CV.

3.4 Relationships among AWC, CT, and CV after Removing Silence

Multiple regressions were performed at each age to explore the relationship among AWC, CT, and CV after the removal of the silence. The results showed that the numbers of AWC and CT could successfully predict the numbers of CV at 10 months, 21 months and 30 months. At 10 months, the results of the regression indicated that the model explained 85.4% of the variance and the model was a significant predictor of the number of CV, $F(2,4) = 18.614$, $p = .009$. While the number of CT contributed significantly to the model (B = 4.194, $p = .004$), the number of AWC did not (B = -.017, $p = .168$). That is, the increase of one unit of CT could contribute to the increase of 4.194 units of CV. At 21 months, the results of the regression indicated that the model explained 91.3% of the variance and that the model was a significant predictor of the number of CV, $F(2,4) = 32.397$, $p = .003$. While the number of CT contributed significantly to the model (B = 3.656, $p = .001$), the number of AWC did not (B = -.054, $p = .058$). That is, the increase of one unit of CT could contribute to the increase of 3.656 units of CV. At 30 months, the results of the regression indicated that the model explained 93.9% of the variance and that the model was a significant predictor of the number of CV, $F(2,4) = 47.429$, $p = .002$. While the number of CT contributed significantly to the model (B = 4.077, $p = .01$), the number of AWC did not (B = -.028, $p = .664$). That is, the increase of one unit of CT could contribute to the increase of 4.077 units of CV. Both sets of analyses indicated that speech directed to children or speech spoken right before or after child vocalizations (i.e. CT) imposed stronger effects to children's vocalizations than speech that was not spoken in temporal proximity to children's vocalizations.

3.5 Cross-language Comparison

With silence time included, the average number of AWC, CT, and CV across the five ages was 634, 14, and 52 per hour (i.e., 634*12 hr=7608, 14*12 hr=168, 52*12 hr=624 per 12-hour day) respectively. Compared with the English normative percentile estimates for AWC, CT, and CV in Gilkerson and Richards (2009), the Chinese-speaking families' AWC in the present study were at the 10^{th}-20^{th} percentile, and CT and CV were below the 10^{th} percentile. With silence excluded, the average number of AWC, CT, and CV across the five ages was 1734, 39, and 150 per hour (20808, 468, 1800 per 12-hour day) respectively. Compared with the English normative percentile estimates for AWC, CT, and CV in Gilkerson and Richards (2009), the

Chinese-speaking families' AWC in the present study were at the 80th-90th percentile, and CV and CT were at the 40th-50th percentile, which were much higher than when silence was included. As discussed earlier, the results with silence excluded were more similar to Zhang *et al.*'s (2015) AWC and CT baseline values; the results with silence excluded can be compared to the results in Gilkerson and Richards (2009). These results showed that the Chinese-speaking caregivers in the present study were on the talkative end of the English normative estimates. However, the Chinese-speaking adults and children were not vocally engaged at similar rates as AWC because the percentile of CT and CV were much lower than percentile of AWC. Gilkerson and Richards (2009) found that children who were first-born, were girls, or had parents with higher education tended to receive more adult talk each day. In the present study, the three factors might have also contributed to high AWC in the present study: 1) All seven mothers were highly educated, having received at least a bachelor's degree, 2) five out of the seven children were first born, and 3) five of the seven children were girls. However, unlike the results reported in Gilkerson and Richards (2009), the talkative caregivers in the present study did not have talkative children.

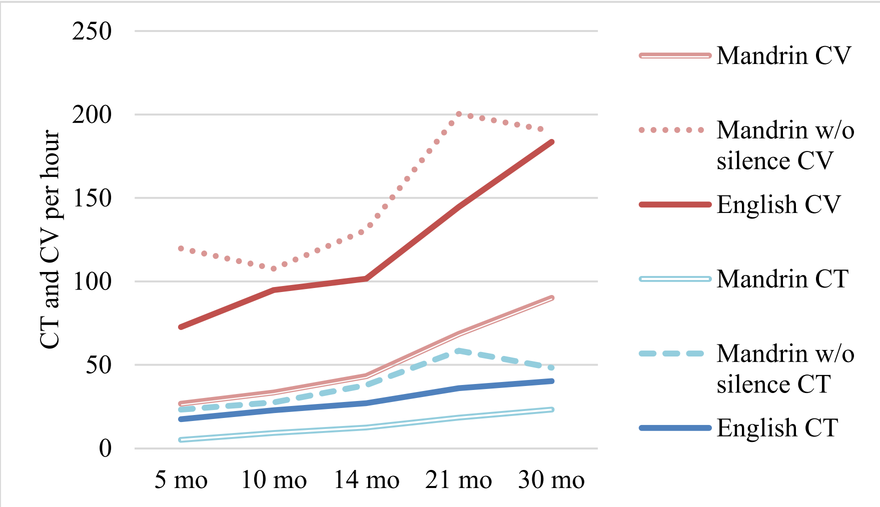

Figure 5. Average adult word (AWC), conversational turn (CT), child vocalization (CV) per hour from the present study and Gilkerson and Richards (2008)

Figure 5 shows longitudinal CT and CV changes in the English-speaking families from Gilkerson and Richards (2008) and the Chinese-speaking families from the present study. Both groups of families showed a gradual increase with age. When silence was included, the Chinese-speaking families showed overall lower CT and CV than the English-speaking families. However, when silence was removed, the Chinese-speaking families showed higher values than the English-speaking families. The group differences could be explained by the fact that the LENA-determined silence not only included times when families were sleeping but also when families were awake but quiet. The results of the two sets of data would be more comparable if the English samples also exclude LENA-determined silence. Another possible reason for the group differences is sample size. More participants and detailed analyses are needed to explore possible cultural differences or confirm the results.

3.6 Limitations and Future Directions

Limitations were identified in the present study and can be addressed in future research. First, a differentiation of the number of child-initiated conversational turns and adult-initiated conversational turns would help examine parent-child interaction patterns and identify the relationship between CT and CV. Now, CT consists of both when a child speaks and an adult responds, and when an adult speaks and the child responds. The LENA Advanced Data Extractor (ADEX, LENA Research Foundation, 2020) would be useful in future research because it provides a more detailed output, including utterances or words of male adults, female adults, the key child, and other children.

Second, to ensure that the key child is really taking turns with another speaker or vice versa, the content of the adult words and child vocalizations requires human coding because the LENA system does not identify the content of the speech sample. For example, it is possible that a parent was holding the key child while talking to another person, but the LENA system may count this parent’s utterances as if she or he were talking to the key child. Third, regarding the unit of speech samples, the LENA system categorizes adult and child speech samples in different units. AWC refers to the number of individual words adults speak, while CV means the number of speech-related utterances produced by the children. When a child produces prelinguistic sounds in a sequence or one breath, the LENA system may count these sounds as one CV. However, when the child starts to produce words or a mixture of babbling and words, the LENA system may still recognize those word strings/vocalizations as one CV. Again, human coding of the recording would be able to identify children’s utterances in word or syllable units when the child starts to produce words.

Furthermore, the results of the present study were only compared with the English normative estimates because Chinese normative estimates using LENA are not available. Developing a Chinese version of the LENA normative estimates would enhance people’s understanding of the effects of early vocal development and adult-child interactions on later development in the Chinese-learning children. Including a larger cohort of participants (i.e., with different socio-economic status, later-born children, male children) to collect a corpus would best represent the Chinese-learning children’s speech capacity at the age.

4. Conclusion

The LENA automated approach has provided researchers with a new recording method that has automatic parsing capacities. The researchers investigated longitudinal changes in the average AWC, CT, and CV with and without silence time, relationship among the three variables, and cross-language comparison in Chinese-learning families with children ranging in age from 5 to 30 months. The percentage of LENA-determined silence decreased with age, indicating that the children's awake time increased as they age. The results also showed that a typically developing Chinese-learning child in the present study listened to an average of 1734 adult words, engaged in 39 conversational turns, and produced 150 vocalizations per hour from 5 to 30 months of age when he or she was awake. Child vocalizations and conversational turns increased over time, but adult word count did not show a clear pattern. When the periods of silence were included, the number of AWC and CT predicted the numbers of CV at 10 months and 30 months. After the periods of silence were removed, the results showed that the numbers of AWC and CT predicted the numbers of CV at 10, 21, and 30 months. This result suggests that the speech produced in temporal proximity to children's vocalizations or directed to children exerted a stronger influence on the number of child vocalizations than the quantity of adult words.

Acknowledgement

This research was supported by Chiang Ching-Kuo Foundation for International Scholarly Exchange in Taiwan to Li-mei Chen for international collaboration with Dr. Kim Oller at the University of Memphis. A special thank you is extended to the families of the children in this longitudinal study for their support of this project.

References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416-430. doi: 10.1111/1467-8624.00287

Ambrose, S. E., VanDam, M., & Moeller, M. P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and Hearing*, *35*(2), 139-147. doi: 10.1097/AUD.0b013e3182a76768

Busch, T., Sangen, A., Vanpoucke, F., & van Wieringen, A. (2018). Correlation and agreement between Language ENvironment Analysis (lenaTM) and manual transcription for Dutch natural language recordings. *Behavior Research Methods*, *50*(5), 1921-1932. doi: 10.3758/s13428-017-0960-0

Bzoch, K. R., League, R., & Brown, V. L. (2003). *Receptive-expressive emergent language test*. Austin, TX : PRO-ED.

Canault, M., Le Normand, M.-T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the Language ENvironment Analysis system (LENATM) in European French. *Behavior Research Methods*, *48*(3), 1109-1124. doi: 10.3758/s13428-015-0634-8

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics*, *128*(5), 910-916. doi: 10.1542/peds.2011-0609

Caskey, M., Stephens, B., Tucker, R., & Vohr, B. (2014). Adult talk in the NICU with preterm infants and developmental outcomes. *Pediatrics*, *133*(3), e578–e584. doi: 10.1542/peds.2013-0104

Charron, C., Fitzpatrick, E. M., McSweeney, E., Rabjohn, K., Somerville, R., & Steacie, P. (2016). Language ENvironment Analysis (LENA) with children with hearing loss: A clinical pilot. *Canadian Journal of Speech-Language Pathology & Audiology*, *40*(1), 93-104.

Fensen, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories*. Baltimore, MD: Paul H. Brookes.

Galland, B. C., Taylor, B. J., Elder, D. E., & Herbison, P. (2012). Normal sleep patterns in infants and children: A systematic review of observational studies. *Sleep Medicine Reviews*, *16*(3), 213-222. doi: 10.1016/j.smrv.2011.06.001

Ganek, H. V., & Eriks-Brophy, A. (2018). A concise protocol for the validation of Language ENvironment Analysis (LENA) conversational turn counts in Vietnamese. *Communication Disorders Quarterly*, *39*(2), 371-380. doi: 10.1177/1525740117705094

Gilkerson, J., & Richards, J. A. (2009). *The power of talk. Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development* (LENA Foundation Technical Report LTR-01-2). Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-01-2_PowerOfTalk.pdf

Gilkerson, J., & Richards, J. A. (2008). *The LENA Natural Language Study* (LENA Foundation Technical Report LTR-02-2). Retrieved from https://www.lena.org/wp-content/uploads/2016/07/LTR-02-2_Natural_Language_Study.pdf

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4), e20174276. doi: 10.1542/peds.2017-4276

Gilkerson, J., Zhang Y., Xu D., Richards J. A., Xu X., Jiang F., …Topping K. (2015). Evaluating language environment analysis system performance for Chinese: A pilot study in Shanghai. *Journal of Speech, Language, and Hearing Research*, *58*(2), 445-452. doi: 10.1044/2015_JSLHR-L-14-0014

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*, 1167. doi: 10.3389/fpsyg.2015.01167

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83-92. doi: 10.1177/1525740110367826

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509-516.doi: 10.1177/0165025406071914

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing.

Ingram, D. (1989). *First Language Acquisition: Method, Description and Explanation*. New York, NY: Cambridge University Press.

Ireton, H. (1992). *Minnesota Child Development Inventory*. Minneapolis, MN: Behavior Science Systems.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., …Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development*, *66*(2), i-149.

Lee, C. C., Jhang, Y., Relyea, G., Chen, L. M., & Oller, D. K. (2018). Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development*, *50*, 140-153. doi: 10.1016/j.infbeh.2017.12.002

LENA Research Foundation. (2020). L*ENA Research Foundation*. Retrieved from http://www.lena.org/

Li, L., Vikani, A. R., Harris, G. C., & Lin, F. R. (2014). Feasibility study to quantify the auditory and social environment of older adults using a digital language processor. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *35*(8), 1301-1305. doi: 10.1097/MAO.0000000000000489

Liu, L., & Kager, R. (2017). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition*, *20*(3), 561-575. doi: 10.1017/S1366728916000183

Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, *20*(1), e12425. doi: 10.1111/desc.12425

Oller, D. K. (2010). All-day recordings to investigate vocabulary development: A case study of a trilingual toddler. *Communication Disorders Quarterly*, *31*(4), 213-222. doi: 10.1177/1525740109358628

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, …Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354-13359. doi: 10.1073/pnas.1003882107

Orena, A. J., Polka, L., & Srouji, J. (2018). Evaluating the LENA recording system for investigating speech input in a French-English bilingual context. *The Journal of the Acoustical Society of America*, *143*(3), 1871-1871. doi: 10.1121/1.5036141

Oshima-Takane, Y. (1988). Children learn from speech not addressed to them: The case of personal pronouns. *Journal of Child Language*, *15*(1), 95-108. doi: 10.1017/S0305000900012071

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child Development*, *67*(2), 621-634. doi: 10.2307/1131836

Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J. A., Ma, L., …Topping, K. (2016). Effects of feedback on parent–child language with infants and toddlers in Korea. *First Language*, *36*(6), 549-569. doi: 10.1177/0142723716649273

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41-58. doi: 10.1017/S0142716400009863

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880-891. doi: 10.1111/desc.12172

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language*, *27*(3), 643-664. doi: 10.1017/S0305000900004232

Richards, J. A., Xu, D., Gilkerson, J., Yapanel U., Gray S., & Paul, T. (2017). Automated assessment of child vocalization development using LENA. *Journal of Speech, Language, and Hearing Research*, *60*(7), 2047-2063. doi: 10.1044/2017_JSLHR-L-16-0157

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762-1774. doi: 10.1111/j.1467-8624.2012.01805.x

Roy, B. C., Frank, M. C., & Roy, D. K. (2009). Exploring word learning in a high-density longitudinal corpus. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, *31*(31), 2106-2111.

Sacks, C., Shay, S., Repplinger, L., Leffel, K. R., Sapolich, S. G., Suskind, E., …Suskind, D. (2013). Pilot testing of a parent-directed intervention (Project ASPIRE) for underserved children who are deaf or hard of hearing. *Child Language Teaching and Therapy*, *30*(1), 91-102. doi: 10.1177/0265659013494873

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672-686. doi: 10.1017/S0305000912000141

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659-673. doi: 10.1111/j.1467-7687.2012.01168.x

Suskind, D., Leffel, K. R., Hernandez, M. W., Sapolich, S. G., Suskind, E., Kirkham, E., …Meehan, P. (2013). An exploratory study of "quantitative linguistic feedback": Effect of LENA feedback on adult language production. *Communication Disorders Quarterly*, *34*(4), 199-209. doi: 10.1177/1525740112473146

Thiemann-Bourque, K. S., Warren, S. F., Brady, N., Gilkerson, J., & Richards, J. A. (2014). Vocal interaction between children with Down Syndrome and their parents. *American Journal of Speech-Language Pathology*, *23*(3), 474-485. doi: 10.1044/2014_AJSLP-12-0010

VanDam, M., Ambrose, S. E., & Moeller, M. P. (2012). Quantity of parental language in the home environments of hard-of-hearing 2-year-olds. *The Journal of Deaf Studies and Deaf Education*, *17*(4), 402-420. doi: 10.1093/deafed/ens025

Velleman, S. L., & Vihman, M. M. (2002). Whole-word phonology and templates: Trap, Bootstrap, or Some of Each? *Language, Speech, and Hearing Services in Schools*, *33*(1), 9-23. doi: 10.1044/0161-1461(2002/002)

Warren, S. F., Gilkerson, J., Richards, J. A., Oller, D. K., Xu, D., Yapanel, U., …Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders*, *40*(5), 555-569. doi: 10.1007/s10803-009-0902-5

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143-2152. doi: 10.1177/0956797613488145

Zhang, Y., Xu, X., Jiang, F., Gilkerson, J., Xu, D., Richards, J. A., … Topping, K. J. (2015). Effects of quantitative linguistic feedback to caregivers of young children: A pilot study in China. *Communication Disorders Quarterly*, *37*(1), 16-24. doi: 10.1177/1525740115575771

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale* [4th ed]. San Antonio, TX: The Psychological Corporation.