INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Parental spoken scaffolding and narrative skills in crowd-sourced storytelling samples of young children

Zhengjun Yue[1], *Jon Barker*[1], *Heidi Christensen*[1], *Cristina McKean*[2], *Elaine Ashton*[2], *Yvonne Wren*[3], *Swapnil Gadgil*[4] *and Rebecca Bright*[4]

[1] Department of Computer Science, University of Sheffield, UK
[2] Education, Communication and Language Sciences, Newcastle University, UK
[3] North Bristol NHS Trust, UK
[4] Therapy Box, London, UK

z.yue@sheffield.ac.uk, rbright@therapy-box.co.uk

Abstract

A novel crowdsourcing project to gather children's storytelling based language samples using a mobile app was undertaken across the United Kingdom. Parents' scaffolding of children's narratives was observed in many of the samples. This study was designed to examine the relationship of scaffolding and young children's narrative language ability in a story retell context which is analysed at the macro-structural (total macro-structure score), the micro-structural (mean length of utterances in morphemes) and verbal productivity (total number of utterances) levels. Young children with and without scaffolding were statistically compared. The interaction between the level of scaffolding support, the grammar complexity and the narrative structure was explored. A bidirectional relationship was observed between scaffolding and young children's narrative language ability. Young children with better performance were observed to receive less scaffolding from parents. Scaffolding was shown to support early narrative development of young children and was more able to benefit those with low-level grammatical complexity skills. It is crucial to encourage parental scaffolding to be well-attuned to the child's narrative ability.

Index Terms: storytelling, children narrative language ability, scaffolding, mean length of utterance in morphemes, narrative macro-structure

1. Introduction

Speech and language therapists (SLTs) sample children's storytelling to support the diagnosis and management of language disorders. This is often a labour-intensive process involving the manual transcription of a child's spoken output and analysis of language complexity. A National Institute for Health Research i4i funded industry, clinical and academic collaboration is designing, developing and evaluating a mobile app: The Language Explorer. This app records a child's storytelling in a controlled context, and uses speech recognition technology to aid in rapid transcription and analysis of the words, phrases, grammar, and story structures the child uses.

Narrative is a form of discourse where the child moves from what is observable to be able to create meaning through language alone [1]. For children, narrative skills progress gradually from sharing information about events with no logical sequence (i.e., in the order that the child wants to relay details), through to telling stories with a clear structure of a beginning, middle and end [2]. Children begin developing narrative skills before they start school, and this skill is a cornerstone for the development of a range of aspects of school readiness [1].

Adults scaffold the skills that children need to complete a task by providing guidance and feedback [3]. At pre-school age, before the child has progressed to learning to read, children's narrative development is supported largely by (parental) adult input. It starts with the provision of information about orientation and continues to support them to develop the story actions and events and finally, narrative evaluation skills [4, 2]. With time, the child internalises this information so that the external scaffold ceases to be needed[5, 6]. Parents need to be responsive to their child's changing abilities in regards to narrative [7]. A recent meta-analysis [8] demonstrated that verbal scaffolding is an effective and widely used method to foster children's early narrative development. Much is still unknown about narrative development and how it varies in context and conversational partners [1]. Narrative skills are of particular difficulty for children with developmental disabilities [9, 10, 11, 12] and therefore it is important to understand the nature of the parental support that promotes the development of narrative skills.

In early trials of the Language Explorer app, an animated and narrated story designed with and for children aged between 4 and 7 years 11 months was presented. Children were asked to retell the story with the pages of the story displayed, and their spoken response was recorded. The story was designed to include opportunities for the child to use a range of grammatical structures and demonstrate their storytelling abilities. As shown in Figure 1, parents were asked to use the app with their children to enable the collection of multiple language samples for training the tool. When listening to the samples, it was observed on multiple occasions that speakers other than the child, assumed to be the parents, could be heard. This was despite instructions within the app to parents to only record the child speaking. Understanding the nature of this verbal scaffolding, its relationship with children's performance in terms of narrative macro-structure, micro-structure elements and verbal productivity, and automating its analysis bring a number of potential benefits. Children with developmental language disabilities often make slow progress with intervention, which is not necessarily detectable through static tests of performance. Dynamic assessments which consider changing performance in relation to the nature of the scaffolding and the need for it over time are thought to be more sensitive measures of children's progress and potential for language development.

In this study, the impact of participant characteristics across age, gender and socioeconomic status on narrative development and scaffolding was analysed for all 600 samples. Then data from children in the sample aged 4;0-4;5 years old were used to investigate narrative skill in a story retell context across the measures of mean length of utterances in morphemes (MLUm), a composite score of the macro-structure of story grammar elements (Total macro-structure (TM) score) and the total number

Copyright © 2021 ISCA

http://dx.doi.org/10.21437/Interspeech.2021-1297

of utterances (TNU). The narrative performance of two groups of children with and without scaffolding was compared to illustrate the bidirectional relationship in an age group expected to benefit from scaffolding. The scaffolding support/response on narrative structure and the interaction between the MLUm and the effects of scaffolding on narrative structure performance were further explored to better understand the nature of the parental support/response to young children's narrative structure development.

Figure 1: *The storytelling task in Language Explorer app.*

2. Materials

2.1. Data collection

A novel crowdsourcing [13] approach was used to collect data via mainstream mobile devices to inform the development of a future clinical tool for children. Compared with the clinic- or lab-based data collection method, crowdsourcing is a more time-effective and feasible process, and it promotes the diversity of the participants [13, 14]. The tool was used to gather samples of typically developing British children aged 4;0 to 7;11. The quality of the audio samples was graded manually to be sufficient for language sampling (transcription and analysis). A stratified sampling model was used to gather a target of 600 samples with stratification by age band, gender and socio-economic status (SES). The SES was determined using a postcode-assigned index of multiple deprivation score (1 to 5) represented in quintiles. The samples were split into 6-month intervals. Since the variation in language is less at 6 and 7 years old, four six month age bands were collapsed into two 12 month age bands. Consequently, six age bands are reported: 4;0-4;5, 4;6-4;11, 5;0-5;5, 5;6-5;11, 6;0-6;11 and 7;0-7;11. The analysed features are automatically generated (e.g., a fine-tuned BERT-based [15] multi-class text classification model was applied for TM score prediction and the Amazon Transcribe with WebSocket [16] was used for speech recognition and speaker separation). The features were then manually improved by Newcastle University speech and language therapy researchers, ensuring the features met an acceptable agreement score (greater than 85%) to move forward to the clinical version of the tool. Table 1 summarises the number of samples.

Table 1: *Number of samples by age band and gender.*

Age band	Male	Female	Total
4;0-4;5	52	49	101
4;6-4;11	51	49	100
5;0-5;5	49	51	100
5;6-5;11	51	49	100
6;0-6;11	51	48	99
7;0-7;11	51	49	100
Total	305	295	600

2.2. Narrative skill measures and scaffolding measures

The MLUm, TM score and TNU were used as measured features of young children's narrative skill in grammatical complexity, narrative structure and verbal productivity in story-retell. MLUm is a micro-structural metric of syntactic complexity which is calculated as the average length of an utterance in morphemes. It has been used as one of the most robust indices of the level of young children's language acquisition and as a valuable measure to diagnose language impairments in young children [17, 18, 19, 20]. The TM score used in the app was adopted from Stein's definition [21], which is the aggregate scoring on seven items marking story grammar - setting, character, initiating event, plan, attempt, consequence and internal response. These features are each given a score out of 3. TNU counts the total number of utterances the child uses in a language sample, as a measure of verbal productivity or talkativeness [7]. The secondary speaker utterance count (SecUC) and secondary speaker word count (SecWC) are measured from the data collected by the app and measure the level of secondary speaker scaffolding. We refer to them as *scaffolding measures.*

3. Results and Discussion

3.1. Participant characteristics effect

We first investigated how participant characteristics, e.g., age, gender and SES, relate to the narrative skill measures and the level of secondary speaker scaffolding. The correlations between age and the five measures are presented in the 5th column of Table 2. Figure 2 plots the mean MLUm, TM, TNU, SecWC and SecUC scores by age band, illustrating how the development of the narrative skill measures and the level of parental scaffolding changes over time. On average, older children demonstrate better performance on measures of MLUm and narrative structure. In contrast, both scaffolding measures have negative associations with age as the amount of parental spoken prompting decreases with time. As also observed in Kelly's study [7], the amount of external scaffold tends to be less required as the child internalises the information learnt in the earlier age. TNU also exhibits a negative correlation with age for children older than five years old.

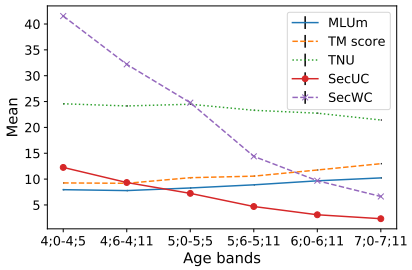

Figure 2: *Narrative and scaffolding measures' means by age.*

A t-test was performed for the two gender groups in regards to the five measures in the 4th column of Table 2. A significantly different performance on TM score and the need of parental scaffolding was observed between the two groups. The 2nd and 3rd columns present the measure means by gender. It is found that, on average, the female group performs better on narrative structure and also receive less parental scaffolding than males. The gender difference is in line with Eriksson's finding [22]. The last column of Table 2 demonstrates that statistically there

Table 2: *Measure statistic by gender, age and SES. * represents that the p-value of Pearson correlation is over 0.05.*

Measures	Male	Female	Gender t-test	Age (r)	SES (r)
MLUm	8.70	8.89	-1.02*	0.41	0.04*
TM score	10.40	10.90	-2.03	0.46	0.05*
TNU	23.03	23.85	-1.38*	-0.16	0.06*
SecUC	7.56	5.49	2.51	-0.33	0.02*
SecWC	25.14	18.01	7.33	-0.31	0.02*

is no correlation of SES on the measures of narrative or the level of parental scaffolding ($p > 0.05$).

To understand how the presence of scaffolding responds to age and gender, we then calculated the proportion of samples without parental scaffolding in Table 3 by age band and gender. When comparing the proportions in 12-month intervals for males and females combined, a smaller proportion of the younger children is shown to complete the narrative task without scaffolding. The presence of scaffolding with children aged 5 and 6 years is similar (31.5% and 32.3%) while a much lower (10.9%) and much higher (54%) proportion without scaffolding is observed for children aged 4 and 7, respectively. The 2nd and 3rd rows reported the proportion for males and females separately. Gender difference exists for children aged 5;6-7;11 years old. The proportion of females that do not receive parental prompting is over 10% higher than that of males. This might because female children cease to receive the external scaffolding earlier than the male children.

Table 3: *Proportion of samples without parental scaffolding by age and gender. M: male, F: female.*

Age band	4;0-4;5	4;6-4;11	5;0-5;5	5;6-5;11	6;0-6;11	7;0-7;11
M+F	11.8%	10%	33%	30%	32.3%	54%
M	10.2%	10.2%	33.3%	22.4%	22.9%	46.9%
F	13.5%	7.8%	32.7%	37.3%	41.2%	60.8%

3.2. Narrative skill measures and scaffolding Correlations

As shown in Figure 2, the 4;0-4;5 age band has the highest level of parental scaffolding. The proportion of samples without parental scaffolding in the 4;0-4;5 age band is the second lowest (only 1.8% over 4;6-4;11 age band). Pre-school children's narrative development is of particular interest among researchers when looking at the nature of parental scaffolding. The following experiments were conducted only on the samples of children aged 4;0-4;5 years old. Figure 3 depicts the correlation matrix (calculated by Pearson correlation coefficient [23]) among narrative skill and scaffolding measures for children aged 4;0-4;5 years old. We selected SecUC as a primary measurement of parental scaffolding level since the SecWC and SecUC could be considered almost perfectly correlated ($r = 0.98$, $p < 0.05$) [24] so that the SecUC could also be used as a proxy for the SecWC. MLUm and TNU are strongly correlated ($|r| > 0.6$) with SecUC. MLUm and TM score have 0.60 and 0.34 negative correlation with SecUC, while TNU has a 0.68 positive correlation with SecUC. This suggests that children with better performance on MLUm and TNU require lower level of parental scaffolding, and more parental spoken prompting seem to motivate the verbal productivity of children. Correlations are also observed between the narrative skill measures themselves (e.g., $r_{MLUm,TMscore}$ = 0.47), indicating that the narrative skill measures are not independently correlated with the level of parental scaffolding.

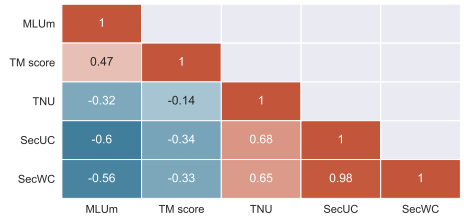

Figure 3: *Correlation matrix (4;0-4;5 age band)*

Table 4: *Statistic difference between the non-parental and parental groups. The p-values are all < 0.05*

	MLUm		TM score		TNU	
	mean	range	mean	range	mean	range
Parental	7.66	2.48-14.45	8.87	2-15	25.10	11-62
Non-parental	9.81	7.56-13.96	11.67	9-16	20.33	15-31
T-test	(-3.38, 0.001)		(-3.28, 0.0014)		0.075	

3.3. Comparing samples with and without scaffolding

To further explore how the presence of parental scaffolding relates to young children's performance on the narrative task, we compared the statistical difference between the two groups: group "parental" and "non-parental", consisting of samples with (SecUC>0) and without (SecUC=0) parent scaffolding in Table 4. A t-test was applied to the three narrative skill measures between the two groups in the last row. It indicates a significant difference between the two groups' narrative performance on MLUm and TM score while the presence of scaffolding does not influence the TNU. The non-parental group is displayed to have higher scores but lower range on MLUm and narrative structure on average compared to the parental group

It is noted that the maximum MLUm a child achieved in the parental group (which had on average poorer performance) is higher than the non-parental group. And the maximum TM score a child in the parental group achieved is only one score less than the non-parental group. This benefit might come from the parental scaffolding support. We checked that the child who achieved the highest TM score (15) in the parental group had 10.12 MLUm. The TM score of the child in the parental group is related to MLUm and scaffolding while it is related to MLUm in the non-parental group. We then checked that the maximum TM score obtained by children with around 10.12 MLUm in the non-parental group is 13, which is less than the score of the child with the highest TM score in parental group. This suggests that with the support of scaffolding, the parental group child scored more highly in regards to narrative structure than the non-parental group child who had similar MLUm with the parental group child. Although the parental group generally performed less well on narrative skills measured than the non-parental group, the scaffolding is capable to support the children in developing their narrative skill.

3.4. Scaffolding on narrative structure

The range of TM scores achieved by children aged 4;0-4;5 years old was from 2 to 16 with a mean of 9.20 (*SD* = 2.91, 95% CIs = 8.62- 9.77). To explore how scaffolding interacts with the children's production of narrative structure, we divided the TM scores into three levels: 0-5, 6-10 and 11-16 and then plotted the SecUC distribution in samples with each TM level in Figure 4. It could be seen that 38% parents prompted more than 20 utterances for children who achieved 0-5 TM scores, while 10%

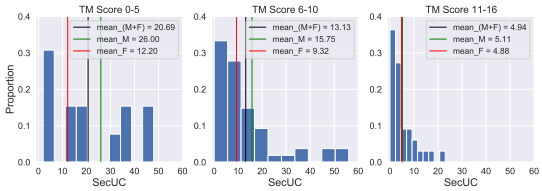

Figure 4: *SecUC distribution in samples with three TM levels.*

and 2% for children who achieved 6-10 and 11-16 TM scores. This indicates that more samples with lower TM scores have larger amounts of scaffolding (SecUC>20). We then calculated the SecUC mean for children with different levels of TM score, as shown by the coloured vertical lines in the plots. Since a significant difference is shown in narrative macro-structure performance between males and females, as explained in Section 3.1, gender is also taken into account here. The black lines are for both males and females, and the green and red lines are for males and females respectively. It is seen that the SecUC mean decreases as the TM level increases, i.e., more scaffolding is observed for samples achieving lower TM scores. Taking gender into account, in each plot, the SecUC mean of the female group is smaller than the male. It is also observed that the gender difference decreases as the macro-structure level increases. This might suggest that the gender difference decreases as the children's narrative ability developed. The male group with lower macro-structure ability received more scaffolding than the females. The female group tended to have better ability to produce narrative independently than the male group.

3.5. MLUm and scaffolding on narrative Interaction

We investigated the interaction between MLUm and effects of scaffolding on narrative structure to learn whether MLUm influences the scaffolding response to the development of young children's narrative ability. A scatterplot that labels the data by TM scores is depicted in Figure 5, which intuitively illustrates the relationship among MLUm, TM score and SecUC, three pairwise related variables. The MLUm of the children aged 4;0-4;5 years old ranges from 2.48 to 14.45, with a mean of 7.91 (*SD* = 2.18, 95% CIs = 7.48 -8.34). It was shown that as MLUm increases, the maximum SecUC decreases. This might be because children with better MLUm performance are more independent and able to complete the narrative task with less spoken prompting from parents.

Children with different MLUms vary in their narrative macro-structure ability. We split the MLUm into three levels: 0-4.8, 4.8-10 and 10-15 to explore the effect of scaffolding support on narrative macro-structure ability. This showed that all samples within the lowest MLUm level (0-4.8) needed the parent's support, while 9.3% and 27.8% samples with 4.8-10 and 10-15 MLUm respectively completed the narrative task without scaffolding. We determined 4.8 as the low MLUm threshold. Figure 5 shows that children with MLUm below the threshold (the red horizontal line) tend to achieve higher TM score when the parent provides more prompting. The highest TM score a child with low MLUm achieved was 6, with more than 50 utterances of parental prompting. However, for children who have similar MLUm above the threshold, more scaffolding does not seem to foster children's narrative ability further. Those who achieved higher TM scores could have a lower level of parent spoken prompting. For instance, for samples with 4.8-7 MLUm, less than 10 SecUC is presented to benefit children's narrative structure. However, additional scaffolding did not appear to

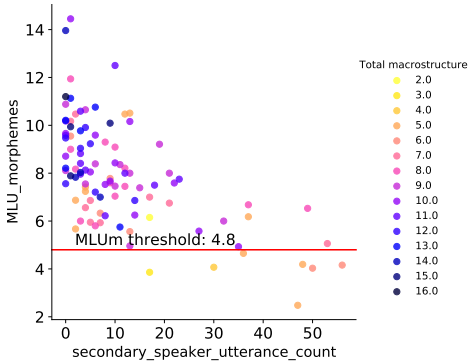

Figure 5: *MLUm and SecUC labeled with TM score.*

help. We then explored the relationship between parental scaffolding and TM score in different MLUm ranges (0-4.8, 4.8-10, 10-15). The maximum TM score for samples in each MLUm level was 6, 15, 16, respectively. The correlations between TM score and SecUC are 0.98, -0.25 and -0.33 respectively. This indicates that the narrative structure ability of children with low MLUm (0-4.8) is more likely to be improved from the support of the parent spoken prompting. This raises the suggestion that parent spoken scaffolding is crucial to help children with low MLUm develop narrative macro-structure skills. More parental scaffolding resulted in higher TM scores. In contrast, for children with higher MLUm, parent prompting was not a case of 'the more, the better'. The right amount of prompting could promote the performance of children with higher MLUm (>4.8). It is wise for parents to provide well-attuned scaffolding which is adjusted to the child's needs and abilities.

4. Conclusions

In this study, a bidirectional relationship between parental scaffolding and children's language ability was observed. Scaffolding is strongly related to children's narrative skills as indicated by MLUm and TM scores. Children with higher narrative ability were shown to require less support from parental spoken scaffolding. Scaffolding could foster children's early narrative development, and may benefit more children with lower MLUm to develop their narrative macro-structure skill. Gender differences were observed in the level of scaffolding received and narrative ability. Males demonstrated lower narrative ability that required more scaffolding than females. The findings could inform the development of future clinical tools for children. This also inspires us to instruct parents to adjust the level of scaffolding to children's needs and abilities. The co-construction of the development of narrative ability and the parental spoken scaffolding is of great interest. Future work will focus on the impact of the time information (e.g., the start time) of parental scaffolding on children's narrative skill development. The nature of scaffolding support (e.g., leading questions, open question or directive statements) and its impact on children's narrative performance could also be explored.

5. Acknowledgements

This work is supported under the EU's H2020 Marie Skłodowska-Curie programme TAPAS (Training Network for PAthological Speech processing; Grant No. 766287). The Language Explorer project is funded by the National Institute for Health Research (NIHR) Innovation for Invention. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

6. References

[1] A. Schick and G. Melzi, "The development of children's oral narratives across contexts," *Early Education and Development*, vol. 21, no. 3, pp. 293–317, 2010.

[2] A. L. Bailey, A. C. Moughamian, K. R. Kelly, A. McCabe, and B. H. Huang, "Leap-frog to literacy: maternal narrative supports differentially relate to child oral language and later reading outcomes," *Early Child Development and Care*, 2018.

[3] L. Vygotsky, E. Hanfmann, and G. Vakar, "Studies in communication. thought and language," 1962.

[4] A. McCabe and P. R. Rollins, "Assessment of preschool narrative skills," *American Journal of Speech-Language Pathology*, vol. 3, no. 1, pp. 45–56, 1994.

[5] A. Collins, J. S. Brown, and S. E. Newman, "Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics," *Thinking: The Journal of Philosophy for Children*, vol. 8, no. 1, pp. 2–10, 1988.

[6] N. Granott, "Scaffolding dynamically toward change: Previous and new perspectives," *New Ideas in Psychology*, vol. 23, no. 3, pp. 140–151, 2005.

[7] K. R. Kelly and A. L. Bailey, "Becoming independent storytellers: Modeling children's development of narrative macrostructure," *First Language*, vol. 33, no. 1, pp. 68–88, 2013.

[8] D. Pesco and A. Gagné, "Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings," *Early Education and Development*, vol. 28, no. 7, pp. 773–793, 2017.

[9] K. R. Biddle, A. McCabe, and L. S. Bliss, "Narrative skills following traumatic brain injury in children and adults," *Journal of communication disorders*, vol. 29, no. 6, pp. 447–469, 1996.

[10] M. Losh and L. Capps, "Narrative ability in high-functioning children with autism or Asperger's syndrome," *Journal of autism and developmental disorders*, vol. 33, no. 3, pp. 239–251, 2003.

[11] J. Reilly, M. Losh, U. Bellugi, and B. Wulfeck, ""Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome," *Brain and language*, vol. 88, no. 2, pp. 229–247, 2004.

[12] A. K. van Bysterveldt, M. F. Westerveld, G. Gillon, and S. Foster-Cohen, "Personal narrative skills of school-aged children with Down syndrome," *International journal of language & communication disorders*, vol. 47, no. 1, pp. 95–105, 2012.

[13] M. Eskenazi, G.-A. Levow, H. Meng, G. Parent, and D. Suendermann, *Crowdsourcing for speech processing: Applications to data collection, transcription and assessment*. John Wiley & Sons, 2013.

[14] R. Bright, S. Gadgil, C. McKean, T. Wren, E. Ashiton, and S. Harding, "Developing a clinical language analysis tool to identify language disorder," in *Child Health Technology Conference*. NIHR, CHT2021.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[16] "Transcribe speech to text in real time using amazon transcribe with websocket," https://aws.amazon.com/blogs/machine-learning/transcribe-speech-to-text-in-real-time-using-amazon-transcribe-with-websocket/, 2019.

[17] R. Wieczorek, "Using MLU to study early language development in english," *Psychology of Language and Communication*, vol. 14, no. 2, pp. 59–69, 2010.

[18] M. L. Rice, F. Smolik, D. Perpich, T. Thompson, N. Rytting, and M. Blossom, "Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments," 2010.

[19] S. L. Eisenberg, T. M. Fersko, and C. Lundgren, "The use of MLU for identifying language impairment in preschool children," 2001.

[20] M.-J. Ezeizabarrena and I. Garcia Fernandez, "Length of utterance, in morphemes or in words?: MLU3-w, a reliable measure of language development in early basque," *Frontiers in psychology*, vol. 8, p. 2265, 2018.

[21] N. L. Stein and C. G. Glenn, "An analysis of story comprehension in elementary school children: A test of a schema." 1975.

[22] M. Eriksson, P. B. Marschik, T. Tulviste, M. Almgren, M. Pérez Pereira, S. Wehberg, L. Marjanovič-Umek, F. Gayraud, M. Kovacevic, and C. Gallego, "Differences between girls and boys in emerging language skills: Evidence from 10 language communities," *British journal of developmental psychology*, vol. 30, no. 2, pp. 326–343, 2012.

[23] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.

[24] M. D. Parker and K. Brorson, "A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw)," *First language*, vol. 25, no. 3, pp. 365–376, 2005.