

# Construction and Analysis of Japanese Parent-Child Dialogic Reading Corpus for Conversational Agents

Yuko Nakagi, Yuya Chiba, Sanae Fujita, Shoko Araki

NTT Communication Science Laboratories

Kyoto Japan

{yuko.nakagi, yuya.chiba, sanae.fujita, shoko.araki}@ntt.com

## Abstract

Dialogic reading, which involves interactive exchanges between a parent and a child during picture book reading, has been shown to effectively promote children's language development. While many support systems for picture book reading have been developed to reduce the burden on parents, existing systems are not yet capable of handling dialogic reading, which requires dynamic parent-child interaction. To develop conversational agents capable of dialogic reading, we constructed a multimodal corpus of parent-child picture-book reading dialogues. The corpus comprises recordings from 36 Japanese parent-child pairs taken during actual picture book reading sessions. In this study, we annotated the corpus with dialogue acts relevant to parent-child communication and categorized the types of quizzes and questions used in the sessions, analyzing the linguistic aspects of parent-child interaction during dialogic reading. After dividing the dialogues into two groups based on the proportion of the child's utterances, our analyses revealed that dialogue systems should adapt their interaction strategies according to individual child characteristics.

**Keywords:** Conversational Systems/Dialogue/Chatbots/Human-Robot Interaction, Corpus (Creation, Annotation, etc.), Text Analytics

## 1. Introduction

Reading picture books together is an important parent-child activity that helps develop children's language and communication skills (Bus et al., 1995; Noble et al., 2019). In particular, *dialogic reading*, in which a parent asks the child about the story and the child shares their own ideas during picture book reading, has been shown to be effective (Whitehurst et al., 1988; Morgan and Meier, 2008; Kotaman, 2008). Nevertheless, due to parental time constraints and particular conditions of home environments, not all families have sufficient opportunities for shared reading.

To address these issues, several systems have been developed to support picture book reading. For example, Xu et al. (2021, 2022a) developed systems that ask prepared questions at predetermined points during reading, while Zhang et al. (2022); Yao et al. (2022) proposed systems that alternate between reading and question-answer modules in a predefined order. Although these systems follow a fixed dialogue sequence, dialogic reading requires systems to manage dynamic parent-child interactions, such as asking and answering questions and sharing impressions during reading.

In this study, to identify parents' dialogue strategies in dialogic reading and then incorporate them in conversational agent design, we constructed a corpus of shared picture book reading involving parent-child dyads. Figure 1 shows an example of the collected data<sup>1</sup>. The recorded interactions



Transcription	
[Parent]	ぐりとぐら。読むよ。 <i>Guri and Gura. I'm going to read now.</i>
[Child]	うん。 OK. (Parent is reading a book.)
[Parent]	大きな何が落ちてた? あ、こっちだった。 <i>What big thing fell? Oh—I mean this one.</i>
[Child]	栗。 <i>A chestnut.</i>
[Parent]	そうだね。栗とこっちは? <i>That's right. A chestnut—and what about this one?</i>
[Child]	どんぐり。 <i>An acorn.</i>
[Parent]	きのこも生えてた。 <i>There were also mushrooms growing.</i> (Parent continues reading.)

Figure 1: Example of parent-child dialogic reading. The italicized transcriptions are translated from the original Japanese.

include audio-visual recordings of parent-child interactions. In total, we collected 70 dialogues from 36 dyads (10.7 hours). We annotated the transcriptions with dialogue acts (DAs) that were relevant to parent-child communication and categorized the quizzes and questions used in the reading sessions. We then conducted linguistic analyses while

<sup>1</sup>The picture book is *Guri and Gura* by Rieko Naka-

gawa and Yuriko Yamawaki (Omura), pp. 24–25.

focusing on how parents scaffold children's speech during interaction.

Our contributions are twofold. First, we constructed a multimodal dialogue corpus comprising spontaneous dialogic reading interactions from 36 parent-child dyads. Second, we conducted linguistic analyses based on the proportion of the child's utterances and demonstrated that parents' interactional strategies and question content varied with the child's speaking characteristics.

## 2. Related Work

### Shared Book Reading and Support Systems

Prior research has demonstrated that shared book reading fosters the development of a wide spectrum of early language abilities in children, including vocabulary (Elley, 1989; Farrant and Zubrick, 2012) and literacy skills (Bus et al., 1995; Justice and Ezell, 2000). In particular, dialogic reading, in which parents prompt children's utterances with open-ended questions and provide contingent feedback as reading proceeds, has been regarded as more effective than simple read-aloud approaches (Whitehurst et al., 1988; Flack et al., 2018).

A variety of systems have been proposed to support picture book reading, including a robot delivering prepared story comments (Michaelis and Mutlu, 2017), conversational agents presenting questions at predetermined moments (Xu et al., 2021, 2022a; Zhang et al., 2022), and question-generation models for picture books (Yao et al., 2022; Xu et al., 2022b). While these studies have reported that their systems can foster story comprehension, they remain incapable of supporting dialogues during picture book reading, which entail the dynamic parent-child interactions observed in dialogic reading.

**Shared Book Reading Corpora** Several corpora have been developed to capture parent-child interactions during shared book reading. Stoops et al. (2024) constructed a corpus comprising 19 hours of home picture book reading by 12 U.S. parent-child dyads; Tare and Gelman (2011) constructed a corpus with audio from 28 bilingual U.S. families reading wordless picture books. Conica et al. (2023) constructed a corpus from audio-video recordings of 46 Irish dyads engaged in diverse picture book reading sessions. For ages 3–7, there is also a corpus of audio-video recordings from 34 U.S. dyads reading picture books (Chen et al., 2020). All of these corpora primarily focus on reading picture books aloud. In contrast, our corpus captures dialogic reading with parent-child questions and includes synchronized audio-video of both speakers.

## 3. Construction of Corpus

### 3.1. Recording of Dialogic Reading

In building our corpus, the interaction was conducted between two separate soundproof booths to record clear speech from both parent and child speakers. This setup permitted the recording of clean audio for each speaker, thus facilitating subsequent speech processing and analysis. Seated in these different rooms, they engaged in a conversation in which the parent read a picture book to the child. Parents were instructed to initiate exchanges about the picture book (e.g., questions about child's impressions and quizzes about the story content) either during the reading or immediately afterward. The conversation was carried out via video chat, with the parent and child viewing each other on laptops in their respective rooms.

Each dyad completed two reading sessions. In the first session, all pairs read the same picture book, a longtime bestseller in Japan that every parent reported being familiar with. In the second session, the child selected one book from five alternatives that differed from the title used in the first session. These candidate books were chosen to closely match the first session's book in total word count and narrative complexity (Fujita et al., 2015). Participants completed questionnaires before and after each session. The pre-session questionnaire assessed reading frequency and book preferences, while the post-session questionnaire asked whether the child liked the book and subsequently felt more motivated to read. All procedures were approved by the ethics committee of NTT Communication Science Laboratories.

The speech was recorded with multiple microphones, and the video was captured using the recording function of the video chat tool. Audio was captured in eight channels: stereo microphones placed in each room, headset microphones for the parent and the child, a lapel microphone on the child's chest, and a condenser microphone placed in the child's room. The recordings were saved as WAV files at 48 kHz, 32-bit float precision, and video was saved in MP4 format.

A total of 36 native Japanese-speaking parent-child dyads participated in the recordings. Among the children, 18 were 3 to 5 years old, 13 were 6 years old, and 5 were 7 years old. Among the parents, 3 were male and 33 were female. Because two dyads completed only the first session, the dataset contains 70 recorded dialogues in total.

### 3.2. Annotation

We annotated the transcripts of the collected dialogues for dialogue acts (DAs) and for the types of quizzes and questions. DAs were labeled at the

Directives	Speech elicitations
Commitments	Declarations
Markings	Statements
Questions	Evaluations
Demands for clarification	Text editing
Vocalizations	Responses
Reading aloud	

Table 1: Dialogue act tags

Overview	
No. of parent-child dyads	36
No. of dialogues	70
Child’s age	3–7
Child’s gender (girl/boy)	16/20
Parent’s gender (female/male)	33/3
Audio	
Total duration (hours)	10.7
Ave. dialogue duration (min.)	9.17 ± 2.15
Utterances	
No. of total child’s utterances	2,699
No. of total parent’s utterances	8,495
Annotations	
No. of total DA tags	11,194
No. of total quizzes	323
No. of total questions	196

Table 2: Corpus statistics

utterance level. Speech was segmented into utterances at every silent interval longer than 0.3 second. Each utterance was automatically transcribed using `whisper-large-v3`<sup>2</sup>, after which errors in voice activity detection and automatic speech recognition were manually corrected.

For DA tagging, we used a 13-category scheme prepared with reference to INCA-A, a label set for parent-child communication (Ninio et al., 1994). Table 1 shows the scheme’s inventory of DA tags. The tags were assigned automatically using `google/gemma-3-27b-it`<sup>3</sup>, an open-source, relatively lightweight model with strong multilingual performance, including handling of Japanese (Kamath et al., 2025). The instructions to the model for DA annotation contained commands to assign the most appropriate tag to each utterance, corresponding to the names and definitions of the 13 DA categories. To evaluate the reliability of the DA tag set, two experts annotated 10 dialogues. Their inter-annotator agreement was Cohen’s  $\kappa = 0.826$ , indicating high consistency. The average agreement between each annotator and the automatic annotation was Cohen’s  $\kappa = 0.591$ , suggesting moderate agreement. In this paper, we regarded the automatic annotation as sufficiently reliable and

<sup>2</sup><https://huggingface.co/openai/whisper-large-v3>

<sup>3</sup><https://huggingface.co/google/gemma-3-27b-it>

F0 scores (ave. ± std.)	
Child (girl)	2.36 ± 0.16
Child (boy)	2.36 ± 0.15
Parent (female)	2.33 ± 0.13
Parent (male)	2.12 ± 0.12
Ave. speech rate (mora/sec)	
Child (girl)	4.19
Child (boy)	4.61
Parent (female)	5.86
Parent (male)	6.16

Table 3: Statistics of prosodic information

Tag name (No. of total tags)	Child (2,699)	Parent (8,495)
Reading aloud	0.0%	50.7%
Declarations	0.1%	5.0%
Statements	11.0%	5.5%
Questions	9.6%	19.0%
Evaluations	3.8%	5.7%
Demands for clarification	8.2%	1.6%
Responses	57.9%	6.3%

Table 4: Proportions of DA tags for parents and children. The Child and Parent columns show the proportion of each DA tag among all child or parent utterances, respectively. Only tags with a proportion of 5% or higher for either speaker are shown.

used it in the subsequent analyses.

We further extracted all prompt utterances produced by the parent and classified them as either *quizzes* or *questions*. We defined a *quiz* as a prompt utterance whose answer can be derived from the content of the picture book, and a *question* as any other prompt utterance (e.g., inquiries about the child’s impressions or about family and friends). Both quizzes and questions were categorized into interrogative form as either WH- or Yes/No-questions. In addition, each quiz was labeled with one of the seven categories from FairytaleQA, a quiz database designed to assess narrative comprehension (Xu et al., 2022b): *Character*, *Causal Relationship*, *Action*, *Setting*, *Feeling*, *Prediction*, and *Outcome Resolution*.

### 3.3. Data Statistics

Table 2 gives the statistics of our corpus. The total duration of the dialogues is 10.7 hours, with a mean of 9.17 min per dialogue. The total numbers of utterances produced by children and parents are 2,699 and 8,495, respectively. We obtained 11,194 DA annotations, and the numbers of quizzes and questions posed by the parent to the child were 323 and 196, respectively.

Table 3 also presents the statistics of prosodic features by speaker. Fundamental frequency (F0)

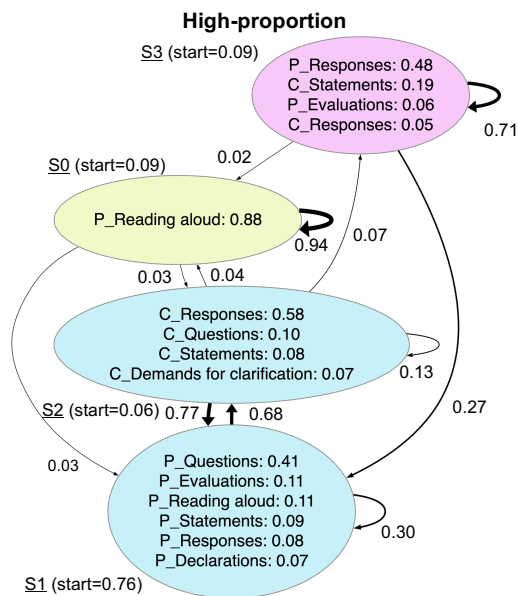


Figure 2: Conversational flows for high-proportion child-utterance groups. Letter before each tag indicates the speaker (P: parent, C: child).

was extracted using PyWORLD<sup>4</sup> and transformed to a logarithmic scale. The F0 values were first averaged within each utterance and then averaged across utterances. Based on the standard deviation of F0, children exhibited greater pitch variability than parents. Moreover, the mean speech rate results indicate that parents slowed their speech in child-directed interaction: Whereas the typical adult conversational speech rate in everyday interactions is approximately 7–8 mora/sec (Maekawa, 2003), the average speech rate of parents in our corpus was around 6 mora/sec (female=5.86, male=6.16 mora/sec). In both parent and child groups, females spoke more slowly than males.

Table 4 shows the proportions of assigned DA tags, separately for parents and children. Parents exhibited higher frequencies of *Reading aloud*, *Questions*, *Responses*, *Evaluations*, *Statements*, and *Declarations*, whereas children showed higher frequencies of *Responses*, *Statements*, *Questions*, and *Demands for clarification*. This distribution indicates that, in dialogic reading, the parent usually takes the lead by reading aloud, asking questions, and giving feedback, while the child answers these questions and adds self-initiated remarks or requests for clarification.

## 4. Analysis of Corpus

To investigate how parent-child dyads interact during dialogic reading, we analyzed the linguistic as-

<sup>4</sup><https://github.com/JeremyCCHsu/Python-Wrapper-for-World-Vocoder>

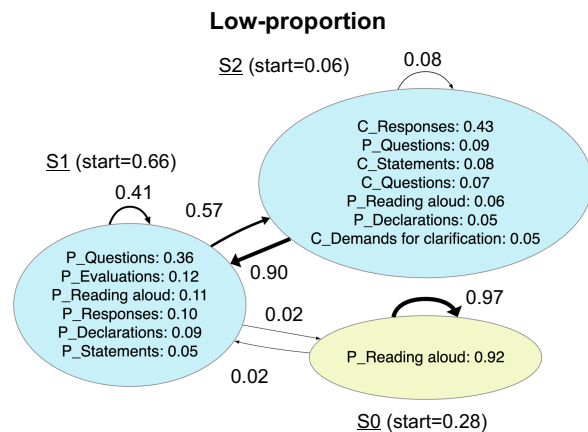


Figure 3: Conversational flows for low-proportion child-utterance groups. Letter before each tag indicates the speaker (P: parent, C: child).

pects of the collected data, focusing on DA tags and types of quizzes and questions. Based on preliminary observations, we found that parent-child interactions could be broadly classified into two types: those in which the child actively participated in the conversation and those in which the child mainly listened. Accordingly, we conducted analyses by dividing the data into two groups based on the quantity of the child’s speech. We calculated the proportion of the child’s utterances for each dialogue and divided the data into high- and low-proportion groups using the median value of 0.23 as a threshold.

## 4.1. Dialogue Acts

### 4.1.1. Modeling of DA Sequences

To analyze conversational flow during dialogic reading, we modeled sequences of DA tags using a hidden Markov model (HMM) (Meguro et al., 2009). We prepared 26 observation symbols for each state, corresponding to the 13 tags for the parent and those for the child. The model was implemented as an ergodic categorical HMM, allowing transitions between all states. We trained the model using the Python package `hmmlearn`<sup>5</sup>. The initial state distribution was set to uniform, and the transition and emission probabilities were randomly initialized. The number of hidden states was varied from 1 to 10 as a hyperparameter. For each configuration, the model was trained for 100 iterations, and the configuration with the lowest Bayesian Information Criterion (BIC) was selected for subsequent analysis.

<sup>5</sup><https://github.com/hmmlearn/hmmlearn>

Role	Text
C	Everything was new to me, but, well, I thought <i>Buruburu</i> would be good. ( <i>Statements</i> )
P	Oh, really? ( <i>Responses</i> )
P	Yeah. ( <i>Responses</i> )
P	I see. ( <i>Responses</i> )
P	Why did you think <i>Buruburu</i> was good? ( <i>Questions</i> )
C	At first, I had my eye on the cat book. ( <i>Statements</i> )
P	Uh-huh. ( <i>Responses</i> )
P	You had your eye on it—yeah, yeah. ( <i>Responses</i> )
C	But when I looked at that story a little, it felt kind of weird... kind of... ( <i>Responses</i> )
P	Mm-hmm. ( <i>Responses</i> )
P	The cat one, right? Uh-huh. ( <i>Responses</i> )
C	So, yeah... because of that... but more than that, um, <i>Buruburu</i> looked more interesting, so I chose <i>Buruburu</i> . ( <i>Statements</i> )
P	Okay. ( <i>Responses</i> )

Table 5: Example of a dialogue segment identified as state S3 based on the most likely state sequence estimation. *Buruburu* refers to picture books. In this segment, the child explains why they want to read *Buruburu*. C and P denote child and parent utterances, respectively. The italicized term at the end of each utterance denotes the relevant DA tag. The transcriptions are translated from the original Japanese.

#### 4.1.2. Analysis of Conversational Flow

Figures 2 and 3 show the HMMs fitted to the DA sequences for each group. To highlight the characteristics of the models, we show only those tag names with emission probabilities greater than 0.05 and transition probabilities greater than 0.02. The edge thickness is proportional to the transition probability. From these figures, we can see that the optimal number of states differs between the two groups. Three states with common characteristics were found in both groups, while the high-proportion group included one additional state.

The three common states can be roughly interpreted as follows: a state in which the parent reads the book (S0), a state in which the parent asks questions or provides evaluations (S1), and a state in which the child gives responses or statements (S2). Examining the transition probabilities, we find that the self-transition probability of the parent's reading state is highest, whereas the transition probabilities between the other two states are also relatively high. These results suggest that the three states represent the fundamental flow of dialogic reading, in which the parent asks questions during book reading and the child responds. In particular, in the

low-proportion group, transitions from the parent's reading state occurred mainly toward the parent's questioning state, indicating that parents tended to lead the conversation.

The state specific to the high-proportion group (S3) represents either the parent's responses to the child's utterances or the child's own statements, and it is characterized by a relatively high self-transition probability. Table 5 presents an example of a dialogue segment that was identified as state S3 based on the most likely state sequence estimation. In this dialogue segment, after ordinary question-answer exchanges, the interaction shifted to the parent's backchannels and follow-up questions, as well as to the child's self-disclosure. This state can thus be interpreted as a phase in which the parent listens attentively to the child. Consequently, the target dialogue system should be capable of not only engaging in simple question-answer exchanges but also employing strategies that encourage the child's utterances and facilitate attentive listening, particularly in conversations where the child actively participates.

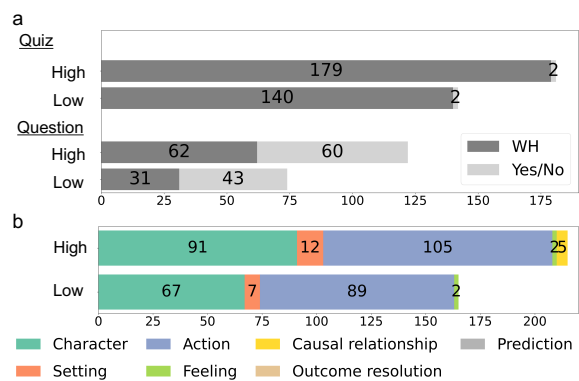


Figure 4: Frequencies of types of quizzes and questions for high- and low-proportion groups

#### 4.2. Types of Quizzes and Questions

Figure 4a shows the numbers of WH- and Yes/No-questions for both quizzes and questions in the high- and low-proportion groups. Quizzes predominantly consist of WH-questions in both groups. However, parents in the high-proportion group tend to use WH-questions more frequently than those in the low-proportion group, suggesting that dialogue systems should adjust the type of prompt utterances to the child's speaking characteristics.

Figure 4b shows the distributions of quiz types by group. From the figures, it can be seen that in both groups, parents mainly used *Character* and *Action* quizzes during dialogic reading. In contrast, *Causal relationships*, *Outcome resolution*, and *Prediction* were rarely observed. These types involve causal or temporal relationships between events

in the story, such as asking why a particular event occurred or predicting what will happen next. Such questions are less likely to appear in dialogic reading, where the story progresses in parallel with the question-answer exchanges.

## 5. Conclusions

With the aim of developing a conversational agent capable of dialogic reading, we constructed a corpus of parent-child picture book reading dialogues. Analyses of dialogue acts and types of quizzes and questions reveal that conversational flow and questioning patterns vary depending on the child's speaking characteristics, suggesting that systems should adapt their interaction strategies accordingly. In future work, we will expand the corpus with additional sessions and conduct fine-grained analyses. For example, we will investigate how children's reading frequency and book preferences influence the flow of dialogue. We also intend to implement our own dialogue system and evaluate its effectiveness through user studies.

## 6. Limitations

The dialogic reading sessions collected in this study are limited to Japanese-speaking parent-child pairs. Consequently, the generalizability of the obtained conversational models across different languages and cultural contexts has not been fully validated. A multimodal English corpus, DAMI-P2C (Chen et al., 2020), captures interactions in which a parent and their child read stories together. Although the nature of these interactions differs in some respects from the dialogic reading data examined in this study, cross-corpus comparisons may help clarify the language dependency of our conversational flow model and contribute to the development of a language-independent model of shared book reading interactions. We plan to conduct such comparative analyses using this corpus in future work.

In addition, a gender imbalance emerged among the parent participants, since no restriction was placed on participant recruitment for the recording sessions. To conduct more reliable gender-specific analyses, it is necessary to increase the amount of data, particularly from male parents.

Furthermore, the present study focused solely on the linguistic aspects of dialogic reading. However, to enhance children's engagement and encourage spontaneous utterances, factors such as tone of voice, speech rate, and turn-taking timing play a crucial role. To develop a multimodal conversational agent, nonverbal behaviors such as facial expressions and gestures are also essential. Since our corpus contains multimodal data capturing interactions between parents and children, we intend

to extend our analysis to these nonverbal behaviors in subsequent studies.

## 7. Ethical Considerations

The data collection was approved by our organization's ethical review committee. Prior to data recording, the study purpose and procedures were explained to the parents, and data were obtained only from parent-child pairs who provided informed consent. Dialogue examples presented in this paper are included only when explicit permission for publication was granted by the parents. To minimize the burden on child participants, all recording equipment was carefully selected and arranged, and sufficient breaks were provided throughout the recording sessions. In addition, to minimize the psychological discomfort of child participants, the recordings of children and parents were made in adjacent booths. While potential self-selection bias may exist due to the nature of the data, all analyses were conducted on anonymized data to ensure objectivity and minimize any bias toward specific individuals or groups.

## 8. Bibliographical References

- Adriana G. Bus, Marinus H. van IJzendoorn, and Anthony D. Pellegrini. 1995. Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, 65(1):1–21.
- Huili Chen, Yue Zhang, Felix Weninger, Rosalind Picard, Cynthia Breazeal, and Hae Won Park. 2020. Dyadic speech-based affect recognition using DAMI-P2C parent-child multimodal interaction dataset. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, page 97–106.
- Mirela Conica, Linda Kelly, Elizabeth Nixon, and Jean Quigley. 2023. Father and toddler language during shared book reading with text-based and wordless picture books. *Read. Res. Q.*, 58(4):655–667.
- Warwick B Elley. 1989. Vocabulary acquisition from listening to stories. *Read. Res. Q.*, 24(2):174.
- Brad M Farrant and Stephen R Zubrick. 2012. Early vocabulary development: The importance of joint attention and parent-child book reading. *First Lang.*, 32(3):343–364.
- Zoe M Flack, Andy P Field, and Jessica S Horst. 2018. The effects of shared storybook reading on

- word learning: A meta-analysis. *Dev. Psychol.*, 54(7):1334–1346.
- Sanae Fujita, Tessei Kobayashi, Yasuhiro Minami, and Hiroaki Sugiyama. 2015. Target age estimation of texts for children. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 22(4):604–620.
- Laura M Justice and Helen K Ezell. 2000. Enhancing children’s print and word awareness through home-based parent intervention. *Am. J. Speech. Lang. Pathol.*, 9(3):257–269.
- Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, and Louis Rouillard et al. 2025. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*.
- Hüseyin Kotaman. 2008. Impacts of dialogical storybook reading on young children’s reading attitudes and vocabulary development. *Reading Improvement*, 43:134–145.
- Kikuo Maekawa. 2003. Corpus of spontaneous Japanese: Its design and evaluation. In *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 7–12.
- Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. 2009. Analysis of listening-oriented dialogue for building listening agents. In *Proceedings of the SIGDIAL 2009 Conference*, pages 124–127.
- Joseph E Michaelis and Bilge Mutlu. 2017. Someone to read with: Design of and experiences with an in-home learning companion robot for reading. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 301–312.
- Paul L. Morgan and Catherine R. Meier. 2008. Dialogic reading’s potential to improve children’s emergent literacy skills and behavior. *Preventing School Failure: Alternative Education for Children and Youth*, 52(4):11–16.
- A Ninio, C E Snow, B A Pan, and P R Rollins. 1994. Classifying communicative acts in children’s interactions. *J. Commun. Disord.*, 27(2):157–187.
- Claire Noble, Giovanni Sala, Michelle Peter, Jamie Lingwood, Caroline Rowland, Fernand Gobet, and Julian Pine. 2019. The impact of shared book reading on children’s language skills: A meta-analysis. *Educational Research Review*, 28:100290.
- Anastasia Stoops, Mengqian Wu, In Ho Ted Jung, and Jessica L Montag. 2024. A novel corpus of naturalistic picture book reading with 2-to-3 year old children. *Language Development Research*, 4(1).
- Medha Tare and Susan A Gelman. 2011. Bilingual parents’ modeling of pragmatic language use in multiparty interactions. *Appl. Psycholinguist.*, 32(4):761–780.
- G J Whitehurst, F L Falco, C J Lonigan, J E Fischel, B D DeBaryshe, M C Valdez-Menchaca, and M Caulfield. 1988. Accelerating language development through picture book reading. *Dev. Psychol.*, 24(4):552–559.
- Ying Xu, Joseph Aubele, Valery Vigil, Andres S. Bustamante, Young-Suk Kim, and Mark Warschauer. 2022a. Dialogue with a conversational agent promotes children’s story comprehension via enhancing engagement. *Child Development*, 93(2):e149–e167.
- Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021. Same benefits, different communication patterns: Comparing children’s reading with a conversational agent vs. a human partner. *Computers Education*, 161:104059.
- Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Jia-Jun Li, Nora Bradford, Branda Sun, Tran Bao Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. 2022b. Fantastic questions and where to find them: FairytaleQA – an authentic dataset for narrative comprehension. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 447–460.
- Bingsheng Yao, Dakuo Wang, Tongshuang Wu, Zheng Zhang, Toby Jia-Jun Li, Mo Yu, and Ying Xu. 2022. It is AI’s turn to ask humans a question: Question-answer pair generation for children’s story books. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 731–744.
- Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. StoryBuddy: A human-AI collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–21.