Wanting to be Understood: Modeling Interaction in Early Language Learning

Qihui Xu* (xu.5430@osu.edu)
Robby Ralston* (ralston.123@osu.edu)
Madison Meares (meares.7@osu.edu)
Vladimir Sloutsky (sloutsky.1@osu.edu)
Department of Psychology, Ohio State University
1835 Neil Ave, Columbus, OH 43210 USA

Abstract

Human language acquisition involves diverse learning resources, including self-supervised learning (sequence prediction) and communicative interactions (talking to caregivers). While recent advancements in language models highlight the power of self-supervised learning, the role of communicative interaction remains unclear. This study uses Reinforcement Learning (RL) and parent-child agent simulations to model interactions and investigate their role in language acquisition, as well as whether RL-like mechanisms may function in children. We pretrained a small transformer model as a child agent, which then interacted with Google's Gemini, acting as a parent agent, to learn language with the goal of being understood. Model evaluations show that the interactive training enhances intelligibility of model's communication and increases behavioral similarity to real child speech. However, minimal pretraining alone provides noticeable syntactic and semantic competence, with RL yielding no consistent gains. These findings imply that interaction may play a more critical role in pragmatic aspects of language learning than in the development of linguistic structures, and that learning through interaction is a mechanism used by children.

Keywords: computational modeling; reinforcement learning; multi-agent interaction; language acquisition

Introduction

During the early years of life, humans rapidly learn to understand and use language, allowing us to think and communicate in ways that are unique across the animal kingdom. Researchers argue that a variety of mechanisms underlie our impressive early language acquisition, including statistical and associative learning (Romberg & Saffran, 2010), self-supervised predictive processing (Liu et al., 2023), mnemonic chunking (Esposito, 2016), rule-like generalization (Endress & Bonatti, 2016), and direct interactions with others (Tomasello & Farrar, 1986). Many researchers also emphasize early, innate constraints on syntax and semantics which reduce the hypothesis space of possible languages and vastly accelerate the learning process (Chomsky, 1980; Markman & Wachtel, 1988; Spelke & Kinzler, 2007).

Like humans, Large Language Models (LLMs) can acquire the ability to proficiently use language in a variety of contexts (Vaswani et al., 2017). However, LLMs are implemented using the domain general transformer architecture and do not have obvious *a priori* constraints on the hypothesis space. Classical models such as GPT-3 are also trained primarily using (self-) supervised predictive learning and forego the other learning mechanisms mentioned in the previous paragraph. Potentially as a result, LLMs require orders of magnitude more data than human children, and receive the equivalent of many hundreds of years of observation before reaching high levels of proficiency (Villalobos et al., 2024). Recent efforts have emphasized this discrepancy, prompting further research and model development that has a more realistic input to better align with human learning (Choshen et al., 2024; Warstadt et al., 2023). Furthermore, even after this lengthy exposure, classical models such as GPT-3 lag substantially behind their human counterparts in semantic and syntactic knowledge (Floridi & Chiriatti, 2020). To achieve higher proficiency, more recent LLMs have been trained with an additional period of Reinforcement Learning with Human Feedback (RLHF), resulting in substantial gains in performance. (OpenAI et al., 2024; Ouyang et al., 2022). During RLHF, a LLM typically generates several responses to a query and receives a human judgment about which response was most suitable. The human judgment can be treated as a reward signal for reinforcement learning, allowing model weights to be fine-tuned to fit user expectations (Chaudhari et al., 2024).

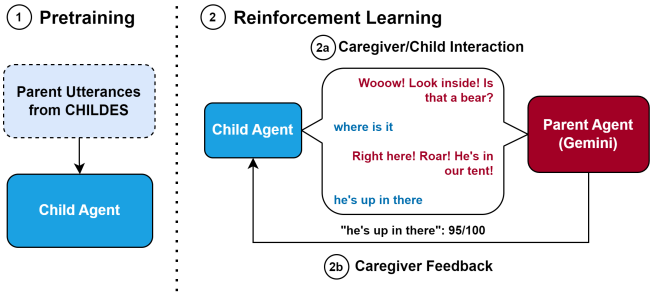

Figure 1: Schematic of our modeling approach.

In this paper, we investigate two key questions: First, whether RLHF can support language acquisition and second, whether the reward-based mechanism underlying RLHF could function as a learning mechanism in children. As a learning mechanism, RLHF requires a) interactions with a human agent, such as a caregiver, and b) sensitivity to the reward signal that the agent wishes to optimize. In humans, young children are intrinsically motivated to pursue linguistic interactions with caregivers and are sensitive to a variety of feedback signals used by caregivers to indicate whether an utter-

*These authors contributed equally to this work.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

ance was (un)acceptable (Goldstein & Schwade, 2008; Nelson, 1973; Nikolaus & Fourtassi, 2023; Zhu et al., 2022). In addition, the quality of child-caregiver interactions is associated with linguistic competence (Cartmill et al., 2013; Clark, 2018, 2020). Together, these suggest that communicative interaction with caregivers could help children learn whether their utterances are appropriate in the current context.

We hypothesize that, after acquiring the basic competence to produce meaningful utterances, children begin to use social feedback, if it is available, to guide further learning (Clark, 2020; Goldstein & Schwade, 2008; Nelson, 1973). This should be viewed as a form of reinforcement learning rather than supervised learning because most cues about intelligibility come from implicit social signals, such as increased engagement from caregivers, rather than explicit corrective feedback (Brown, 1970; Saxton, 1997; Schoneberger, 2010). We hypothesize that a reward signal is inferred from a caregiver's verbal and nonverbal behavior which differs following intelligible vs. unintelligible utterances (Demetras et al., 1986; Kuhl, 2007; Warlaumont et al., 2014).

Such a hypothesis has been difficult to test with traditional approaches due to the challenges of explicitly quantifying rewards and the complexity of manipulating or tracking caregiver interactions with language learners. Previous modeling work (Nikolaus & Fourtassi, 2021) has used RL; however, these studies were conducted in non-interactive settings with simplified reward metrics (e.g., the BLEU score; (Papineni et al., 2002)). With recent advancements in computational modeling and the demonstrated potential of RL (OpenAI et al., 2024), we now have the tools to model these dynamics.

To test our hypothesis (see Fig.1 for schematic), we first pretrained a small transformer model on parent utterances from the CHILDES database, compiled from real parent-child interactions (henceforth the *pretrained model*). We then constructed a series of artificial interactions between the "child agent" (the *interactive model*) and Google's Gemini, a fully-trained, modern LLM, instructed to respond like a parent (Anil et al., 2023). During the interactions, we had the "parent" model both respond to the child agent and give a score for the intelligibility of the child agent's previous response[1]. Intelligibility scores were then used as a reward signal for RLHF using a typical Proximal Policy Optimization (PPO) procedure (Christiano et al., 2017; Schulman et al., 2017). After generating the score, the child model responded to the parent's utterance, allowing the procedure to be iterated into a conversation. By evaluating and comparing performance across different aspects of linguistic competence (i.e., syntax, semantics, and communicative intelligibility), we investigate whether and how the learning mechanisms underlying RLHF support language acquisition. By comparing model behaviors to real child behaviors, we assess whether RLHF functions as a learning mechanism in humans—specifically, whether leveraging social feedback from caregivers, aimed at achieving understanding, plays an important role in language development.

[1]In this study, we use an intelligibility score as a simplified metric of communicative success. While real-world parent–child interactions rely on implicit feedback (e.g., confirmations, clarifications), this approach provides a systematic way to assess whether the intended message is conveyed

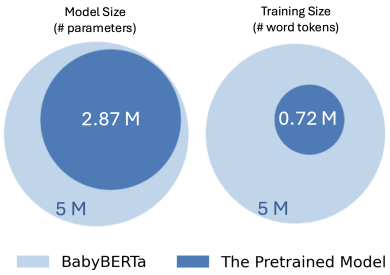

Figure 2: Size comparison between BabyBERTa (a baby language model) and our pretrained model.

Methods

The Child Agent

We used the GPT-2 backbone (Radford et al., 2019; Wolf et al., 2020), pretrained from scratch on a subset of the North American CHILDES Corpus (MacWhinney, 2000) containing parents' speech to infants up to 12 months old. The training corpus contains 724,516 word tokens, 9,911 word types, and 184,918 utterances. We lowercased all text and removed punctuation except for apostrophes (e.g., "'s"). To model utterance transitions, we concatenated every two utterances with beginning- and end-of-sentence symbols. We concatenated utterances to get a call and response or parent and child utterance. We randomly split the dataset into 90% for training and 10% for validation. A tokenizer was trained on the entire training data to reflect infant-directed vocabulary. The model's objective was to predict individual tokens within sequences.

Importantly, this model is substantially smaller even than common baby models like BabyBERTa (Huebner et al., 2021) (Fig.2 shows the size contrast). Hyperparameters were tuned using Optuna with a TPE (Tree-structured Parzen Estimator) sampler (Akiba et al., 2019), which adaptively targets promising regions to minimize validation loss. After 100 trials, the best GPT-2 model had two hidden layers, eight attention heads, a 256-dimensional embedding, a 512-dimensional feed-forward block, a batch size of 32, and an Adam learning rate of 0.0013, totaling 2,867,712 parameters.

The Parent Agent

We used the Google Gemini-pro-1.5 API [2] (Anil et al., 2023) as our "parent" agent for its strong performance in generating coherent, context-sensitive language and its cost-effectiveness for experimentation.

Parent-child interactions lasted 25 conversational turns. The parent model was first given a description of a play scenario (the "context") and instructed to begin the conversation.

[2]Data collection took place from November 24 to 26, 2024.

After each turn, it was prompted to produce the next parent utterance and to evaluate the child's previous response. Parent utterances were obtained using a "parent prompt" and evaluation was done via a "score prompt." We finally cued the child agent with the new utterance to complete the turn. Conversations were built up over time, allowing us to observe how the models interacted during each conversation.

Parent prompts began with, "You are an average parent speaking to their [age]-month-old child," and also included instructions, a description of the context, a summary, and a script of recent utterances. The age began at 12 months, the age immediately after the end of our pretraining data, and increased linearly through training until 36 months, the age associated with our validation set (see below). The agent was instructed to actively listen, encourage verbal responses, and maintain a playful tone. Two sample interactions from the CHILDES corpus illustrated typical parent–child exchanges. Utterances were added sequentially to the script. On turns 5/15 or 10/20 (counterbalanced) we prompted Gemini to summarize the conversation, replaced the existing summary, and cleared the script.

The play context was designed to simulate realistic interactions between parents and children aged 9 to 40 months. We created around 200 parent-child interaction scenarios, each following a structured framework detailing the time, location, activity, and objects involved. For example, the Sensory Exploration scenario features a mid-morning play session where a parent sets up a sensory bin with textured objects like fabric squares, water beads, and plastic spoons, describing each texture to the child.

The second prompt, the "score prompt," began with "You are a teacher observing a conversation between a parent and a 12-month-old child." We framed Gemini as a teacher in this context to avoid score inflation. Gemini rated intelligibility of the most recent utterance on a 0–100 scale. High ratings indicate that the utterance is "understandable because it directly expresses a specific request, making its meaning readily apparent." Conversely, a low rating suggests that the utterance "requires additional effort to understand the context and intention behind the request." An utterance does not need to be syntactically or semantically well-formed to be rated as intelligible. Each range of the intelligibility scale is accompanied by detailed descriptions (e.g., 90-100 for utterances highly specific, easily understandable responses; 0 for non-responses or incomprehensible nonwords). Repetitive utterances are penalized to encourage varied, meaningful outputs. To ensure stable evaluations, we set the temperature to 0 for the score prompt, while a temperature of 0.8 for the parent prompt encouraged more diverse responses.

Convergent validity of our Gemini-based intelligibility measure was evaluated by randomly selecting 125 utterances generated by the interactive model and obtaining scores from GPT-4 (OpenAI, 2023). GPT-4 showed a moderate-to-strong correlation with Gemini ($r = 0.57, p < .001$).

Modeling Interaction through Reinforcement Learning

As described above, every utterance from the child model was assigned an intelligibility score by the parent agent. We treated intelligibility scores as a reward signal for reinforcement learning. Consistent with large-scale applications (Ouyang et al., 2022), we used a Proximal Policy Optimization (PPO) algorithm to find weights that maximize the expected long-term reward. PPO is a 'conservative' policy gradient method, aiming to maximize rewards but also prevent large updates in the wrong direction (Lillicrap et al., 2015). There are three components of the loss function: the policy gradient component favoring actions that produce larger rewards, a KL-divergence term which prevents too-large steps, and an entropy term favoring exploration. Tunable hyperparameters of PPO include coefficients for each component of the loss function, the learning rate for the Adam optimizer, and an initialization value for the KL divergence threshold (Schulman et al., 2017).

To implement PPO, we used a customized version of the Transformer Reinforcement Learning (trl) package in Python, to which we added an explicit entropy coefficient as in Schulman et al. (2017). To simulate young children's limited memory capacity, we used a small batch size (five utterances), and one gradient accumulation step. Otherwise, parameters were left at their default values. Unlike common implementations, we updated model weights as soon as a batch was available, and continued the interaction with the updated weights. We used Optuna's TPE to obtain values for the five hyperparameters that minimize validation loss on a separate dataset (child-directed speech to children aged 35-36 months), which captures an age range at the very end of the reinforcement learning window. After 13 tuning trials—each running for 500 epochs—we observed a clear pattern leading to the optimal parameter set of 10^{-6} for learning rate, 0.01 for entropy coefficient, and 0.1 for value function coefficient, policy gradient coefficient, and KL threshold.

Evaluation Metrics

The evaluation metrics encompass multiple aspects of language competence and the model's predictive fit to real child speech. For syntactic and semantic evaluations, we analyzed the models' internal representations and, to align with our goal of understanding early language learning environments, specifically selected child-plausible tasks if possible.

Intelligibility We used Gemini-generated intelligibility scores[3] to assess the interactive model's ability to produce clear, contextually relevant utterances (see 'The Parent Agent' for details). Intelligibility was recorded each epoch as part of our training procedure.

[3]While evaluation metrics typically go beyond the reward signal used during training, Gemini's judgments, based on trillions of word tokens from human interactions, are arguably aligned with our training goals better than other obtainable measures. We therefore report these scores for evaluation, though we are exploring other options for future testing.

Capturing immaturity We evaluated the similarity of model-generated utterances to real child speech from the Providence corpus in CHILDES (Demuth et al., 2006) under two conditions. First, we compared the model's utterances when talking to Gemini to actual child utterances over training. To do this, we collected the interactive model's output every 100 training epochs and compared it to real child speech. Second, we assessed whether the model's responses to real parent utterances became more child-like over training. For this, we examined the child agent every 500 epochs of RL training, prompted each with real parent utterances, and compared the resulting responses to real child responses.

For real child and parent speech, we selected child utterances at 20 months and parent utterances recorded after 12 months from Providence (Demuth et al., 2006) to avoid overlap with our training data. We used two methods for measuring similarity: BERT sentence embeddings (with mean pooling; Reimers and Gurevych, 2019) and Gemini text embeddings (text-embedding-004; Google AI, 2024). BERT provides widely adopted embeddings, while Gemini offers a state-of-the-art approach, allowing us to cross-validate the results.

Syntactic measures

Syntactic category representation Syntactic category representation aims to examine the models' internal organization of word categories, such as nouns, verbs, and adjectives. To quantify these representations, we employed Representational Similarity Analysis (RSA; Kriegeskorte, Mur, and Bandettini, 2008) to assess both the pretrained and interactive models.

Words for RSA were selected based on their frequency within the training set. Using the spaCy English model ("en_core_web_sm"; Explosion AI, 2023), we first performed part-of-speech (POS) tagging to identify tokens corresponding to key syntactic categories: determiners, nouns, verbs, pronouns, and adjectives. For each category, we extracted POS tags and counted the frequency of each word, selecting the most frequent 100 words as targets.

We next constructed similarity matrices separately for the child model and the ground-truth categories. For the child model, we used pairwise cosine similarities between word embeddings in the second (final) hidden layer. To obtain word embeddings, we presented the model with all sentences containing the target word from the training data and averaged over the obtained representations. We also tested embeddings from the first layer, though results were the same. For the ground-truth categories, the similarity matrix was binary, with a similarity score of 1 assigned to word pairs from the same syntactic category and 0 for pairs from different categories. Finally, we used Spearman's rank correlation to compare the similarity structure of the model's syntactic category representations with that of the ground-truth categories.

WUG test We constructed an evaluation based on the Wug Test, a classic linguistic task measuring children's ability to apply grammatical rules to novel words and demonstrating their capacity to generalize beyond memorized examples (Gleason, 1958). In this task, the model was presented with sentences containing made-up words (e.g., "wug") and asked to apply known grammatical rules, such as pluralization (e.g., "wugs"). Gleason found that children as young as four successfully apply morphological rules to novel words in this task, revealing the generalization of linguistic rules beyond surface-level memorization.

For our evaluation, we created two sentences for each of eleven novel words: a base sequence that used the root form, and a modified sequence that altered the word to follow a grammatical rule base sequences used singular nouns and present-tense verbs (e.g., "wug" and "spow") and modified sequences used compound words and suffixes (e.g., "wug-house", "wugs", and "spowed"). The correct choice in each pair was the modified sequence, as it aligns with the expected grammatical structure given the context. For example, in the sequence "This is a wug. Now there is another one. There are two [wug/wugs]," the plural form is the grammatically correct choice. To present novel words to the child model, we manually tokenized each word to ensure that the same root token appeared across sequences. For instance, the sequence "this is a wug now there is another one there are two wugs" was constructed by adding the suffix "s" at the end of the tokenized input of the word "wug".

Both the pretrained and interactive models were evaluated using the Wug Test, treating the sequence with a lower perplexity score as the model's response. Perplexity is the inverse probability of the sequence, normalized by the number of words. Lower perplexity indicates that the model finds the sequence more predictable, and thus more consistent with its current linguistic understanding. After obtaining perplexity, we counted the sentence-pairs where the model correctly selected the modified sequence over the base sequence.

Semantic measures

Semantic category representation Similar to the approach for syntactic categories, we employed RSA (Kriegeskorte, Mur, & Bandettini, 2008) to assess the models' internal representations to ground-truth semantic categories. We first selected words from the MacArthur-Bates Communicative Development Inventories (MCDI) available in Wordbank (English (American) Words and Gestures; Frank et al., 2017), a database that provides normative data on early language development, including semantic categories for words commonly used by young children, such as animals (e.g., dog) and food and drink (e.g., apple). We constructed similarity matrices based on pairwise cosine similarities of the model-generated embeddings for these words. Word embeddings were obtained by averaging the embeddings of each word across sentences in which it appeared. Finally, we compared the model's similarity matrix with the ground-truth semantic category matrix using Spearman's rank correlation.

Odd one out test The Odd-One-Out test evaluated the model's ability to distinguish semantic discordance within triads of words. For real children, this task is used to assess seman-

tic knowledge in both typical and atypical development, as well as categorization and visual-spatial memory (Kotov et al., 2024; Ruiz, 2011). Each triad consisted of two semantically related words and one "odd-one-out". For instance, in the triad "mommy", "banana", and "baby","banana" would be the correct response. To generate the triads, we used OpenAI's GPT-4 (OpenAI, 2023) with a tailored prompt requesting 100 odd-one-out word triads suitable for 12-month-olds. Model output was verified to ensure words were age-appropriate and representative of a 12-month-old's linguistic environment.

For each triad, we obtained word representations from the second hidden layer of the child model, following the method described above. For each word, we calculated pairwise cosine similarities to both other words in the triad and took the mean. The word with the lowest average similarity was selected as the model's response. Accuracy was evaluated by counting the number of triads where the model correctly identified the odd-one-out.

Results

Syntactic/semantic Competence from Pretraining

For the pretrained model, we first evaluated syntactic competence. Using RSA, we found that word-word similarity derived from model representations correlated with "ground-truth" similarity from syntactic categories, with Spearman's rank correlation of $\rho = 0.35$, $p < .001$, 95% CI $[0.34, 0.36]$ (Fig.3a). However, on the Wug test, the model correctly inflected only 5 out of 11 items, performing numerically below chance level. This suggests that pretraining allowed the model to acquire some syntactic distinctions, though it exhibited limited morphological generalization to novel words. Semantic competence was also reflected in word representations (Fig.3b); word-word similarity correlated with ground-truth semantic categories, $\rho = 0.26$, $p < .001$, 95% CI $[0.23, 0.30]$. In the odd-one-out task, the model achieved an accuracy of 51.02%, exceeding the baseline of 33.3%, showing that its word representations distinguish semantic relationships beyond chance level.

The interactive model did not yield consistent improvements. There was no evidence for change in syntactic organization, as assessed by RSA ($z = -0.00, p = .498$) with $\rho = 0.35$, $p < .001$, 95% CI $[0.33, 0.36]$. Wug test performance slightly declined to 4 out of 11 items, numerically below chance level. Semantic competence also showed a numerical decrease, though it was not statistically significant ($z = 0.41, p = .340$), with $r_s = 0.22$, $p < .001$, 95% CI $[0.20, 0.25]$. However, odd-one-out accuracy marginally improved to 52.04%, numerically higher than the pretrained model.

Overall, these analyses revealed that pretraining alone equips the model with the beginnings of syntactic and semantic competence. This is impressive considering the model's limited depth, relatively few parameters and small training set (Fig.2). In addition, there was no evidence of consistent syntactic or semantic gains from RL. We therefore turned to other metrics to investigate the role of RL.

RL Improved Communication Intelligibility

A regression analysis showed increased intelligibility scores over training (Fig.3c; see also script examples with low and high intelligibility scores), $b = 0.003$, $t(3999) = 21.08$, $p < .001$, $R^2 = 0.10$. This suggests that RL may enhance communication intelligibility. We return to this finding in the Discussion section.

RL Increased Similarity to Real Child Speech

Over the course of RL training, the interactive model's output became increasingly similar to actual child utterances, whether assessed with BERT embeddings, $b = 1.22 \times 10^{-5}$, $t(48) = 11.71$, $p < .001$, $R^2 = 0.74$, or Gemini embeddings, $b = 1.34 \times 10^{-6}$, $t(48) = 10.01$, $p < .001$, $R^2 = 0.68$. Furthermore, when the child agent was configured to respond to real parent utterances, the similarity between its output and child responses to the same utterances improved through training, whether measured with BERT embeddings, $b = 4.07 \times 10^{-6}$, $t(9) = 9.16$, $p < .001$, $R^2 = 0.90$, or Gemini embeddings, $b = 1.08 \times 10^{-6}$, $t(9) = 3.71$, $p = .005$, $R^2 = 0.61$. This suggests that RL makes the interactive model's production more child-like, both in general language production and in contextually grounded interactions (Fig.3d).

Discussion

This study investigated how wanting to be understood in caregiver-child interactions could affect language learning. We used RL to quantify goal-driven learning and simulated parent–child communicative dynamics using agent-based models, enabling flexible, adaptive language behaviors beyond passive input-driven learning. Our key findings reveal that interaction-driven learning significantly enhances communicative intelligibility and increases the model's behavioral similarity to real child speech. However, even minimal pretraining focused on next-token prediction equips the model with the beginnings of syntactic and semantic competence, with no consistent additional gains from interaction. These simulation findings offer several insights into the mechanisms of language learning.

First, the improvement in communicative intelligibility in the interactive model suggests that the drive to be understood can serve as a learning signal, even without explicit instruction. This highlights a fundamental possibility: intelligibility can emerge from optimizing for communicative success, without requiring advanced linguistic competence. In other words, learners could learn to be effectively understood by simplifying utterances, choosing contextually appropriate words, or leveraging pragmatic cues, even when their underlying grammatical knowledge is still developing. This aligns with observations in early child language, where simple, clear utterances often precede more complex linguistic structures (Brown, 1973; Xu et al., 2023).

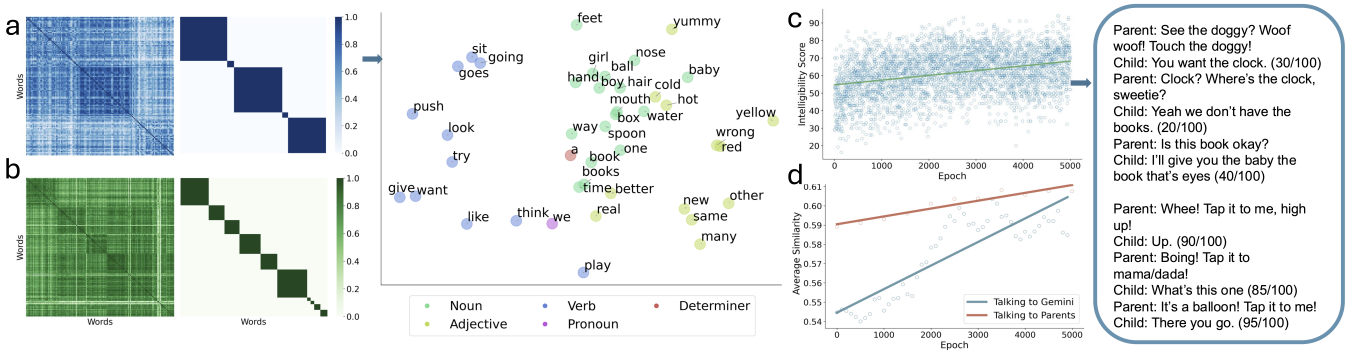

Figure 3: **(a)** Representational Similarity Analysis (RSA) of syntactic categories. We show the similarity matrix for the pretrained model's representations (left) and ground-truth syntactic categories (right). The scatterplot shows model representations using t-SNE. **(b)** RSA for semantic categories using MCDI words. **(c)** Intelligibility scores for the interactive model's generated utterances across training epochs, with a regression fit line. Two script examples are provided (right): one from an early training epoch with a lower intelligibility score and one from a later training epoch with a higher intelligibility score. **(d)** Similarity of the interactive model's generated utterances to real child speech when interacting with Gemini and responding to real parent utterances.

Interestingly, RL not only improved intelligibility but also led the model to produce utterances that more closely resembled real child speech. This suggests that striving to be understood may naturally push language learners toward developmentally typical, "immature" speech patterns. In this sense, the model learns to "be a child," where limited knowledge makes it adaptive to prioritize simplicity, aligning with the Gricean maxim of quantity (Grice, 1975), which encourages providing just enough information to be understood. This strategic simplicity increases communicative success when linguistic competence is still developing, supporting theories that early child language is shaped by the dual pressures of limited cognitive resources and the need to communicate effectively (Piaget, 1955).

Second, self-supervised learning from linguistic input is effective in supporting the development of core linguistic structures. The pretrained model demonstrated noticeable internal organization of syntactic and semantic categories for its size, and distinguished semantic relationships in the odd-one-out test that sometimes elude even older children (Gleason, 1958). However, these competencies did not translate into successful morphological generalization (e.g., the Wug test). These findings align with theories of language acquisition that emphasize the power of distributional learning (Landauer & Dumais, 1997), where exposure to language patterns is effective for acquiring foundational linguistic knowledge.

On the other hand, RL and interactive feedback showed limited benefits for models in acquiring structural aspects of language. This suggests that adapting behavior with the goal of being understood can improved certain pragmatic abilities, such as communicative intelligibility, but it does not necessarily enhance underlying linguistic representations beyond self-supervised learning. This raises important questions about the conditions under which interactive learning can meaningfully support structural language development. Notably, this study focused solely on the goal of being understood, and factors beyond intelligibility may be needed for further growth. In real-world communication, a child's ultimate goal is often to fulfill personal needs (e.g., wanting water), with being understood as an intermediate step. Future work should explore whether more complex, intrinsic motivations to better capture the richness of human language learning.

Future work could improve and extend the current approach in several additional ways. Besides incorporating more complex reward signals and goals, future research should systematically validate developmentally-relevant evaluation metrics, such as intelligibility scores and behavioral similarity to real child language, against human judgements for clarity, naturalness, and developmental appropriateness. Additionally, exploring the impact of pretrained model size—both architecture and input data—on gains from interaction could reveal whether greater knowledge reduces simplicity, diverging from child-like patterns. Future work could also integrate self-supervised statistical learning during the interaction process to systematically compare how much each learning mechanism benefits from interaction. Finally, while the current RL model captures child-like behavior as an outcome, future research should investigate mechanisms to simulate the developmental trajectory of child behaviors, offering a more dynamic view of language and communication across time.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*, 2623–2633.

Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.-B., et al. (2023). Gemini: A family of highly capable multimodal models.

Brown, R. (1970). Derivational complexity and order of acquisition. *Cognition and Development of Language*.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Chaudhari, S., Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., Deshpande, A., & da Silva, B. C. (2024). Rlhf deciphered: A critical analysis of reinforcement learning from human feedback for llms.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Choshen, L., Cotterell, R., Hu, M. Y., Linzen, T., Mueller, A., Ross, C., Warstadt, A., Wilcox, E., Williams, A., & Zhuang, C. (2024). Call for papers: The 2nd babylm challenge: Sample-efficient pretraining on a developmentally plausible corpus.

Christiano, P. F., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems (NeurIPS)*, 4299–4307.

Clark, E. (2018). Conversation and language acquisition: A pragmatic approach. *Language Learning and Development*, *14*(3), 170–185.

Clark, E. (2020). Conversational repair and the acquisition of language. *Discourse Processes*, *57*(5–6), 441–459.

Demetras, M. J., Post, K. N., & Snow, C. E. (1986). Feedback to first language learners: The role of repetitions and clarification questions. *Journal of child language*, *13*(2), 275–292.

Demuth, K., Culbertson, J., & Alter, J. (2006). The providence corpus [Accessed: 2024-02-01].

Endress, A., & Bonatti, L. (2016). Words, rules, and mechanisms of language acquisition. *WIREs Cognitive Science*, *7*, 19–35.

Esposito, J. (2016). Mnemonics as a cognitive-linguistic network of meaningful relationships. *The Journal of Language Learning and Teaching*, *6*.

Explosion AI. (2023). Spacy 'en_core_web_sm' model [Version 3.x, Accessed: 2025-02-01].

Floridi, L., & Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681–694.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data [Accessed: 2024-02-01].

Gleason, J. B. (1958). The child's learning of english morphology. *WORD*, *14*, 150–177.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*, 515–523.

Google AI. (2024). Embeddings in the gemini api [Accessed: 2025-02-01].

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, volume 3: Speech acts* (pp. 41–58). Academic Press.

Huebner, P. A., Sulem, E., Fisher, C., & Roth, D. (2021). Babyberta: Learning more grammar with small-scale child-directed language. *Proceedings of the 25th Conference on Computational Natural Language Learning*, 624–646.

Kotov, A., Ashlan, I., & Kotova, T. (2024). Find (and remember) the odd one out: The effect of categorical distinctiveness on recognition memory. *Collabra: Psychology*, *10*.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4), 1–28.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*(1), 110–120.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Liu, O., Tang, H., & Goldwater, S. (2023). Self-supervised predictive coding models encode speaker and phonetic information in orthogonal subspaces.

MacWhinney, B. (2000). The childes project: Tools for analyzing talk [Accessed: 2024-02-01].

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psycholgy*, *20*, 121–157.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*, 1–135.

Nikolaus, M., & Fourtassi, A. (2023). Communicative feedback in language acquisition. *New Ideas in Psychology*, *68*, 100985.

Nikolaus, M., & Fourtassi, A. (2021, November). Modeling the interaction between perception-based and production-based learning in children's early acquisition of semantic knowledge. In A. Bisazza & O. Abend (Eds.), *Proceedings of the 25th conference on computational natural language learning* (pp. 391–407). Association for Computational Linguistics.

OpenAI. (2023). Chatgpt-4: Large language model [Accessed: 2025-02-01].

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J.,

Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024). Gpt-4 technical report.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, *35*, 27730–27744.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002, July). Bleu: A method for automatic evaluation of machine translation. In P. Isabelle, E. Charniak, & D. Lin (Eds.), *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 311–318). Association for Computational Linguistics.

Piaget, J. (1955). *The language and thought of the child* [Originally published in French in 1923]. World Publishing Company.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Better language models and their implications.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3982–3992.

Romberg, A. R., & Saffran, J. R. (2010). Satistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 906–914.

Ruiz, P. (2011). Building and solving odd-one-out classification problems: A systematic approach. *Intelligence*, *39*, 342–350.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language*, *24*(1), 139–161.

Schoneberger, T. (2010). Three myths from the language acquisition literature. *The Analysis of Verbal Behavior*, *26*(1), 107–131.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *56*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems 30*.

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024, 21–27 Jul). Position: Will we run out of data? Limits of LLM scaling based on human-generated data. In R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, & F. Berkenkamp (Eds.), *Proceedings of the 41st international conference on machine learning* (pp. 49523–49544, Vol. 235). PMLR.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychological science*, *25*(7), 1314–1324.

Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjape, B., Williams, A., Linzen, T., & Cotterell, R. (2023). Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora. *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, 1–34.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.

Xu, Q., Chodorow, M., & Valian, V. (2023). How infants' utterances grow: A probabilistic account of early language development. *Cognition*, *230*, 105275.

Zhu, H., Bisk, Y., & Neubig, G. (2022). Language learning from communicative goals and linguistic input. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, *44*, 1351–1358.