Large Language Models and Children Have Different Learning Trajectories in Determiner Acquisition

Olivia La Fiandra*
University of Maryland, College Park
olivia.lafiandra@gmail.com

Nathalie Fernandez Echeverri*
University of California, Berkeley
nfe@berkeley.edu

Patrick Shafto
Rutgers University, Newark
patrick.shafto@gmail.com

Naomi H. Feldman
University of Maryland, College Park
nhf@umd.edu

Abstract

Large language models are often compared to human learners based on the amount of training data required or the end state capabilities of a learner, yet less attention has been given to differences in their language learning *process*. This study uses determiner acquisition as a case study to characterize how LLMs and children differ in their learning processes. By analyzing annotated speech samples from specified age ranges of four children and intermediate training checkpoints of the Pythia-70m language model, we trace the learners' learning paths of definite and indefinite determiner use. Our results reveal a divergence: the children first produce the indefinite determiner, while the model first produces the definite determiner. This difference reflects underlying differences in the learning goals and mechanisms of models and children. Framing language learning as movement over distributions of linguistic features makes the learning process visible and offers an alternative approach for comparing humans and language models.

1 Introduction

Researchers have often looked to human language learning to quantify progress in language modeling. However, most of the existing evidence for these comparisons comes from two sources: sample efficiency and end-state benchmarks. Sample efficiency measures the linguistic input required to learn language. Large language models (LLMs) are considerably less sample efficient than human language learners. While a child of age 12 will hear less than 100 million tokens in their language environment, language models are trained on data containing billions to trillions of tokens (Warstadt et al., 2023). To compare how language models' learning compares to that of human learners, researchers often rely on benchmarks that characterize the end state of the model. For example, LLMs now achieve high accuracy on evaluating grammatical well-formedness (Papadimitriou et al., 2022), yet they still show weakness in handling certain syntactic dependencies compared to humans (Marvin and Linzen, 2018). Additionally, some models fail to generalize grammatical knowledge to novel contexts that require knowledge of structural relationships (e.g. the relationship between the subject and object of a verb) (Wilson et al., 2023). Although these approaches clearly demonstrate that models are different from humans, they provide little insight into what is different in the *learning process*. Examining the learning process itself could reveal possible disparities between models and humans such as whether linguistic knowledge is acquired under different initial conditions, at different speeds, in different orders, or with varying consistency. How best to quantify these differences remains an open question.

In this paper, we take a new approach to characterizing the differences in learning between LLMs and humans, by looking at the learning trajectory for the acquisition of determiners. Specifically, we examine the definite article *the* and the indefinite article *a* that occur before a noun to specify its referent.

While prior work on language models has focused on sample efficiency and end-state benchmarks, comparing model behavior to child language acquisition provides a complementary perspective on models' determiner acquisition. Some elicitation studies on children's determiner acquisition suggest that children overuse *the* in indefinite contexts (Wexler, 2011; Maratsos, 1976). However, other developmental research shows that children acquire singular definite determiners like *the* rapidly and in adult-like ways in naturalistic production from as early as 1.5–2 years of age (Ying et al., 2024). These findings suggest that children's determiner acquisition is guided by both linguistic input and emerging pragmatic competence, high-

*These authors contributed equally to this work.

lighting the importance of examining learning trajectories rather than just end-state performance. Including child data allows us to situate model learning in relation to human acquisition and provides a benchmark for evaluating not only what models learn, but also how learning unfolds over time.

In our data, we find that children first produce the indefinite article, whereas the model we test first produces the definite article. This difference is not only about which forms are produced, but also about the order and pattern of acquisition over time. By analyzing these trajectories, we can see that models and children may prioritize different aspects of language and acquire determiners in different sequences. We argue that this divergence reflects fundamental differences in how LLMs and children approach language learning, offering insight into the mechanisms underlying the language learning process.

2 Methods

To investigate the way language models build their linguistic knowledge, we use determiner acquisition as a case study. We define determiner use as a multinomial distribution where for each determiner phrase produced, one of three events can occur: a definite article like *the* is used, an indefinite article like *a* or *an* is used, or the required determiner is omitted.

We annotated samples of both children's and a model's determiner use, detailed in Sections 2.1 and 2.2, to trace each learner's learning trajectory. We outline the annotation processes for the child and model data in Section 2.3. We define a learning trajectory as the distributional shifts in determiner usage over time. For example, a learner might initially omit all determiners and then progress to a roughly equal distribution of definite and indefinite determiners. The learning trajectory captures this shift from the initial to the final distribution by tracing distributions of determiner use throughout the acquisition process.

To visualize a learning trajectory, we plot points representing the learner's determiner use on a simplex which maps three points on a 2D triangular plane. Each trajectory begins with the learner's initial distribution of determiner use and progresses toward a defined target distribution, specified separately for the child and model data in Sections 2.1 and 2.2, respectively. Figure 1 illustrates the trajectory of a learner who starts with an equal use of the definite, indefinite, and omitted determiners and gradually shifts toward a balanced distribution of definite and indefinite determiners.

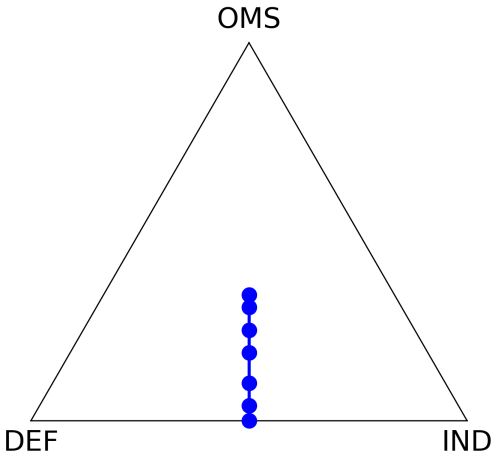

Figure 1: Learning trajectory moving from one-third event distribution to equal distribution of definite and indefinite determiners.

2.1 Child Data

We sampled the child data from the Braunwald and Providence Eng-NA CHILDES corpora (Braunwald, 1997; Demuth et al., 2006; Fernandez et al., 2024). These samples were taken from speech between children and adults, so the children's input is adult speech. We annotated cases where a required determiner was omitted in samples of four children's speech throughout early childhood. For Child 1, we annotated a sample of 4443 lines that spanned across the child's early childhood from 18 months-old to 40 months-old. For the other three children, we annotated 6 samples of their speech, one sample for each age range. For these three children's samples, we aimed to annotate 100 determiner uses per sample, although, in some samples, fewer than 100 determiner uses occurred. We also annotated samples of each child's parent's speech. We sampled and annotated the parent's speech in the same way as we did the children's speech. To check the reliability of the annotations, two annotators independently annotated a sample of the data, yielding a Kappa score of 0.99. One of the two annotators, whose reliability we measured, annotated the child data.

After we annotated the data, we used the determiner counts to plot the children's learning trajectories. We used the children's determiner distri-

butions at the 18.0–22.0 age range as their initial determiner distributions. We defined the children's target distributions based on the distributions calculated from their parents' determiner use.

2.2 Model Data

The model used throughout this study is the Pythia-70m model from the Pythia Suite (Biderman et al., 2023). We chose to use a Pythia model because it contains 154 intermediate training checkpoints. This allows us to probe the model's language use throughout training. Pythia-70m is an autoregressive causal model trained to predict the next token given all previous tokens in the context. Pythia-70m was trained on the Pile dataset. The model saw approximately 300 billion tokens in total, and each intermediate checkpoint of the model processed a batch of approximately 2 billion tokens. These checkpoints include a checkpoint at every 1000 training steps from 0 to 143000. Additionally, the checkpoints include 10 log-spaced checkpoints ranging from 1 to 512. We sampled linguistic output from the following checkpoints: 128, 256, 512, 1000, 2000, 3000, 4000, 5000, 10000, 17000, 35000, 53000, 71000, 107000, and 143000.

To parallel the input children received, we prompted the model with adult speech. Just as children hear and respond to adult speech, the model's input consisted of adult utterances. For every checkpoint we tested, the model received the same 100 lines sampled from a parent's speech in the Braunwald corpus. These lines were randomly selected and contained more than three words. After each prompt, the model generated a response of up to 20 tokens using a greedy-decoding strategy, selecting the token with the highest probability at each step. Each response included a repetition of the prompt followed by the newly generated tokens. This design ensures that the differences we observe between child and model trajectories reflect the learners' processes rather than characteristics of the input.

After counting the uses of determiners at each checkpoint, we calculated the learning trajectory. We used the determiner distribution at checkpoint 128 as the initial determiner distribution. We chose this as the initial distribution because checkpoint 128 was the earliest checkpoint to consistently produce language. The target distribution used to determine the learning trajectory of the model was the model's distribution of determiner use after training. This is the distribution found at the final checkpoint, checkpoint 143000.

2.3 Annotations

To find the learning trajectories of the model and children, we annotated the samples to find the learners' productions of definite, indefinite, and omitted determiners. This study tracked the definite determiner *the* and the indefinite determiners *a* and *an*. We did not track other determiners like *these* and *which*.

We annotated the children's determiners in the stem and gloss fields included in the CHAT format of the corpora. The stem field corresponds to the base form of the utterances, while the gloss field corresponds to the utterances' intended meanings. We annotated determiner omissions by marking the omission with a *0* preceding the omitted determiner in the stem field and the determiner occurring in the gloss field. We annotated appropriate determiner uses with the determiner occurring in both the stem and gloss fields. Table 1 illustrates examples of these annotations.

Next, we annotated the model's output for determiner use, tracking definite, indefinite, and omitted determiners across its linguistic output. Because the model's production was restricted to 20 tokens, it occasionally cut off its response on a determiner. This type of determiner use was annotated as an End of Response Use. Additionally, the model occasionally cut itself off and started a new line of its response. In the cases where the model cut itself off on a determiner and started a new line of its response, we annotated these determiners as Cut Off Uses. We counted End of Response Uses and Cut Off Uses toward the total number of determiners produced by the model in order to accurately calculate the distributions of determiner uses. This is because these incomplete productions still indicate the model's choice to use a determiner in that context, and excluding them would underestimate the determiner frequency. Because the model first repeated the prompt and then generated new tokens in its response, the model may repeat determiners from the prompt in its response. We did not count these determiner occurrences as determiner productions by the model. Table 2 contains examples of these annotations.

For both the model and the children, each consecutive use of the same determiner counted toward the total definite and indefinite uses. For example,

Type	Use	Stem	Gloss
Definite	I saw the dog	I saw the dog	I saw the dog
Indefinite	She is not a toy	She is not a toy	She is not a toy
Omission	It has book inside it	It has 0a book inside it	It has a book inside it

Table 1: Examples of annotations for the child data.

Prompt	Response	Definite	Indefinite	Omission	End of Response	Cut Off
It's a dangerous toy if you can't abide by the rules	It's a dangerous toy if you can't abide by the rules. "I'm not going to be a toy that I'm not going to be a"	0	2	0	1	0
Okay here you go	Okay here you go to the "I'm not sure what you're going to do," he said. "	1	0	0	0	1

Table 2: Examples of annotations for the model data.

the the ball counts as two definite determiner uses.

3 Learning Trajectories

Figure 2 shows the learning trajectory of the model and the average learning trajectory of the children, accompanied by the children's individual trajectories for reference. We assume that before any learning occurs, children would produce a 100% omission distribution, because without having learned determiners, they would never use one in a required context. This represents an unobserved portion of the trajectory, with a hypothetical point at 100% omissions preceding our first measurement. We do not assume the same hypothetical point for the model, and instead expand on the model's pre-learning distribution in Section 4.

At the first measured point, the model exhibits a 100% definite, 0% indefinite, 0% omission distribution. Over the course of training, the model shifts toward a more balanced distribution of the definite and indefinite determiners while never producing omissions. The model's trajectory indicates that the model first produces the definite determiner before the indefinite determiner. In contrast, the children's initial distributions favor the indefinite determiner over the definite, showing that they produce the indefinite determiner first. These differences in initial determiner production reflect that the model and children acquire determiners in different sequences, and this difference in order of acquisition indicates that their learning processes operate differently.

Despite these different sequences of acquisition, both the model and children converge near approximately equal distributions of definite and indefinite determiners by the end of their trajectories. This convergence suggests that despite differing orders of acquisition, both the children and the model ultimately reach a similar endpoint in determiner use. This highlights that different learning processes can produce comparable outcomes, and that the learning trajectory may be an informative measure of language learning.

One way to interpret the comparison is to focus on the portion of the trajectory after children start leaning toward determiner production instead of omission. From this perspective, the model and children behave similarly in how they approach the target distributions, and this can be quantified by measuring their *distance* from the target at each point in time. This measure captures how far a learner's determiner use is from the expected distribution and allows us to examine how quickly that distance decreases relative to the amount of observed data. Figure 3 plots the children's distances and the average distance to the target at each age range and shows the model's distance to the target at each checkpoint on a log scale. The distances were calculated using KL divergence. These figures illustrate how the gap between learner output and the target narrows over time in both the model and children, while also showing that the model's trajectory accelerates more quickly over the course of learning than the children's. Examining whether the trajectory accelerates or slows at different points provides a way to compare learning patterns.

The behavior of the model first producing the definite determiner and then the indefinite determiner can potentially be explained based on the model's language objectives. The objective of

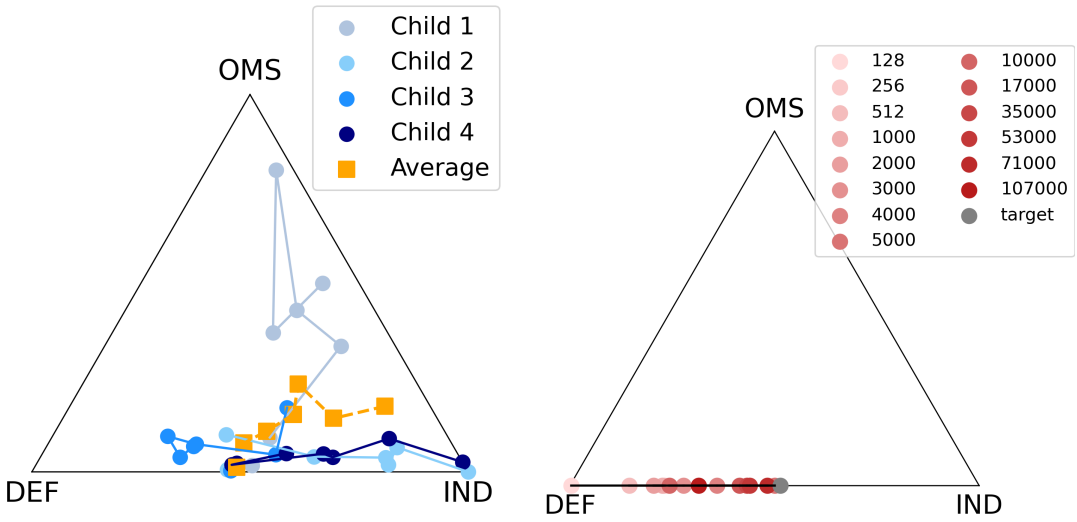

Figure 2: Children's individual and average trajectories (left) and model's learning trajectory (right).

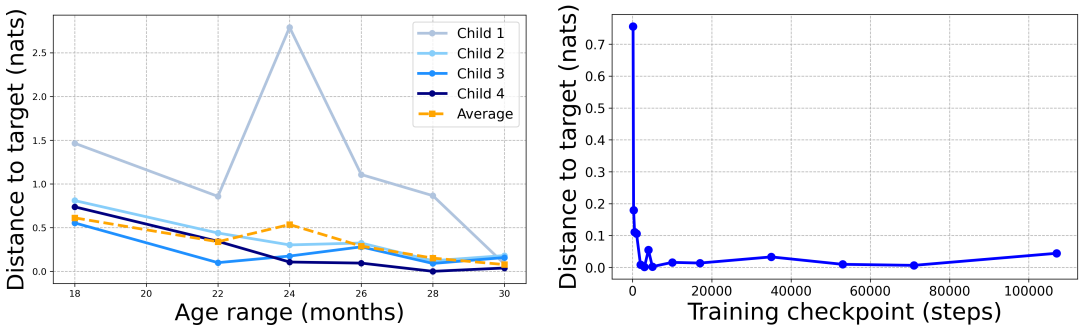

Figure 3: Children's individual and average distances from target at each age range (left) and model's distance from target at each training checkpoint.

Pythia-70m is to generate language by predicting the next word based on the linguistic context. To do this, the model generates the word with a high probability of occurring next based on the prior linguistic context. In early stages of training, the model may generate the definite determiner more often than the indefinite determiner due to the high frequency of the definite determiner occurring in language. For example, data from the Corpus of Contemporary American English (COCA) show that *the* occurs roughly twice as often as *a* or *an* (Davies, 2008). It may be the case that, when learning from the frequencies in its input, the model may initially gravitate toward the more frequent option before it has enough exposure to produce the indefinite determiner.

What this does not explain is why the model does not probability match by the end of training. We should not expect to see the model reach an equal distribution of definite and indefinite determiners at the end of training if the definite determiner is more likely to occur than the indefinite determiner in its training data. Though this is puzzling, it is not inconsistent with the children's target distributions which also approximate equal distributions of the definite and indefinite determiners. This pattern suggests that factors other than data frequencies may contribute to shaping the model's final determiner distribution.

While the model's initial behavior aligns with the fact that the frequency of the definite determiner is higher than the frequency of the indefinite determiner in language input, the children's behavior does not. Despite the definite determiner occurring twice as often as the indefinite determiner in language input, children appear to first produce the indefinite determiner before they produce the definite determiner. One possible reason for this could be that it is easier for children to grasp the meaning of the indefinite determiner than the definite determiner. The meaning of the definite determiner is connected to concepts of uniqueness and

specificity. There is a difference between saying *I have the ball* and *I have a ball*. *The ball* refers to a specific object that is salient to the speaker. It may be the case that children do not yet understand these concepts. Supporting this, a study on infants' perspective-taking in language comprehension found that 14-month-olds do not demonstrate an understanding of specificity, while 19-month-olds do (Choi et al., 2018). In this experiment, two agents and a participant interacted with an apparatus containing two identical balls. The participant and Agent 1 could see both of the balls while Agent 2 could only see one of the balls. When Agent 2 requested *the ball* from Agent 1, 19-month-olds expected Agent 1 to hand over the specific ball visible to Agent 2, showing sensitivity to the uniqueness and saliency of the ball. In contrast, 14-month-olds accepted either ball as a valid referent. This developmental gap between 14-month-olds and 19-month-olds suggests that younger children have not yet grasped concepts of uniqueness and specificity that are required to appropriately use the definite determiner. A lack of understanding of uniqueness, despite a high frequency of the definite determiner in language input, offers a possible explanation for why children's production patterns diverge from frequency-based expectations.

4 Qualitative Analysis

To better understand the model's early behavior, we conducted a qualitative analysis of its determiner use at early training checkpoints. This approach allows us to examine unexpected patterns that quantitative measures might not capture. One such pattern is the model's 100% definite distribution at the first checkpoint we measured.

To investigate the model's frequent production of the definite determiner in early stages of training, we turned to examine the content of the model's responses at checkpoint 128. We found that the model never appropriately uses either determiner despite frequently producing the definite determiner. This is because after repeating the prompt, the model looped its responses ending with determiners. Table 3 demonstrates some of these responses.

The model's behavior at checkpoint 128 leads us to ask why the model repeats the determiner and loops phrases like *and the* that end in a determiner. One possible explanation for this behavior is that the words that the model loops are the only words the model has learned at that point. This would explain why the model only generates the same words—specifically the words *and* and *the*. This explanation also suggests that these are the first words the model learns throughout training. Under this explanation, the model's distribution of determiner use prior to our observing its behavior may not exist. If the first words the model produces are *the* and *and*, then there is never a time when the model omits a required determiner before determiner-learning begins. This is because language learning for the model begins with learning the definite determiner. While the model behaves this way early on in training, we do not see the same behavior from the children. This qualitative difference again suggests that the model and children learn determiners in different ways.

5 General Discussion

This study provides a comparison of determiner acquisition between children and a large language model, highlighting how learning trajectories reveal differences in the sequencing of language learning. Our results show that the model and children differ primarily in where each begins its trajectory: the model initially exhibits a strong preference for the definite determiner, whereas children start with a higher proportion of indefinite determiner use. Over time, both learners converge toward a roughly balanced distribution of definite and indefinite determiners. These findings go beyond prior work by examining not just the end state of language learning or the amount of linguistic input needed to learn, but also the learning process itself, illustrating how sequences of acquisition can differ across humans and models.

While we assume that children begin with a relatively high level of omissions of determiners and learn toward the indefinite determiner once determiner-learning begins, we find that the model begins with 100% use of the definite determiners. These differing starting points raise questions about the underlying mechanisms driving early behavior. Two possible ideas explored in Sections 5.1 and 5.2 could account for this divergence.

In addition to these differences in starting points, a factor shaping the model's learning trajectory is the decoding strategy used. We used greedy decoding, the model's default decoding strategy, which deterministically selects the most probable token at each step. This approach may exaggerate early pref-

Prompt	Response
okay here you go	okay here you go, and the, and the the the the the the the the the the the the the the the the
we're recording our voices	we're recording our voices and the the the the the the the the the the the the the the the the the the the the the the
I did wanna hear why don't you sing it together	I did wanna hear why don't you sing it together, and the, and the, and the, and the, and the, and the, and

Table 3: Examples of responses at checkpoint 128.

erences, such as the model's strong initial bias toward the definite determiner. Alternative decoding strategies could have produced more variable determiner distributions, potentially smoothing or delaying the observed trajectory. While our findings show how determiner use unfolds under greedy decoding, future work should explore whether other decoding strategies alter the model's trajectory or reveal additional stages of learning.

Another observation from our findings is the disconnect between the child data and prior developmental work noted in Section 1. While our results show that children's productions include relatively frequent use of the indefinite determiner early on, prior developmental research has demonstrated children's tendency to overuse definite determiners in experimental contexts. One way to reconcile this discrepancy is to consider differences in experimental and naturalistic contexts. Experimental tasks may introduce pressures that favor definite determiners, whereas naturalistic production data, like the data analyzed here, capture children's baseline use more directly. This comparison emphasizes the importance of context in shaping conclusions about early determiner use and highlights the value of including child data in our comparison with the model. The child data not only illustrates how the model diverges from human learners, but also reveals how different methodological perspectives can yield distinct views of children's developmental trajectory.

5.1 Children and Models Are Different Types of Learners

One possible explanation for the difference between the model and children's learning processes is that children and models are different types of language learners. The model used in this study generates language by predicting the next word based on the probability of the word occurring in a specific context. The model forms these probabilities based on the frequencies of words in its training data, per its training objective. Because the definite determiner occurs frequently in English, the model may have a preference for learning it before learning the indefinite determiner. This would explain why the model starts its learning trajectory with a higher probability of using the definite determiner than using the indefinite determiner. In contrast, these same frequency distributions in children's linguistic environments may not be sufficient on their own to make it easier for children to produce the definite determiner, given underlying conceptual challenges involved in its appropriate use (Arunachalam and Waxman, 2010; Booth et al., 2005). One such conceptual challenge may be understanding the concepts of uniqueness and specificity, discussed in Section 3. If it is the case that language learning is impacted by conceptual challenges, then children may not be able to overcome the challenge of appropriately using the definite determiner over the indefinite determiner using frequencies alone.

5.2 Children and Models Use Language Differently

Another explanation for the difference between learning trajectories of the model and children is that language models and children use language differently. The objective of our language model is to generate language by predicting the next word. In order to complete this goal, the model chooses the word with a high probability of occurring next in the sentence. The goal of a child is different from the goal of a language model. A child's goal is not to predict the next word. Rather, their goal is to communicate (Tomasello, 2003). One possible explanation which would require further research

is that children may tolerate certain errors, including omissions of determiners, in order to efficiently communicate. For example, research has found dissociation between production and comprehension in language use, suggesting that speakers sometimes produce ungrammatical utterances because they cannot retrieve the appropriate form in the moment (Harmon and Kapatsinski, 2017). This could explain why the children in our study omit determiners. The meaning of a child's message is likely not affected by the omission of a determiner, so children may tolerate those mistakes depending on other production challenges they face. For example, a child asking for their bottle may correctly say *I want the bottle* or incorrectly say *I want bottle*. In this case, though omitting a determiner is ungrammatical, the omission does not impact the meaning of the message the child is attempting to communicate.

6 Conclusion

This study examined differences in language learning between models and children by exploring their learning trajectories. We found that models and children behave differently throughout determiner acquisition in regards to their learning processes. The model and the children appear to produce the definite and indefinite determiners in opposite order: the model beginning with the definite determiner and the children beginning with the indefinite determiner. These learning differences reflect disparity in the goals and learning processes that shape language models and humans. It appears that next-word prediction objectives and probabilistic optimization drive determiner use for the model, while communicative needs and learning milestones drive determiner use for children.

The approach of framing acquisition as movement over distributions makes the underlying process of acquisition visible and provides a nuanced basis for comparing language models to human learners. Learning trajectories provide a way of measuring *how* the learning process unfolds for different learners. This study shows a difference in the sequencing of determiner acquisition in models and children, and additional data could reveal further differences such as differences in speed or consistency. Future work could also extend this approach to other aspects of grammar and work to build a broader map of learning differences between models and humans.

By comparing models and humans through their learning trajectories rather than end state learning or the quantity of training data, this approach uncovers fundamental differences in how LLMs and humans learn language over time. Comparing learning processes makes visible the paths and intermediate steps of language learning which offers insights into the mechanisms driving the learning process. This perspective emphasizes the importance of studying *how* learning unfolds, not just *what* is learned or *how much data* is needed. Ultimately, this process-oriented approach provides an alternative way to evaluate and improve LLMs.

Limitations

While this study provides insights into the differences between LLMs and children's determiner acquisition processes, several limitations should be noted. First, the analysis is based on a single large language model and a relatively small set of child learners. The findings may not generalize across other models with different architectures or optimization objectives, nor across child learners with varying linguistic backgrounds. Second, the study focuses exclusively on the articles *the*, *a*, and *an* which represent only a subset of English determiners. Examining a broader range of determiners could reveal different learning trajectories. Third, we measured the model's behavior through its productions, and did not probe its logits or embeddings. This limits our ability to interpret aspects of the model's behavior, such as repeatedly producing the definite determiner. While inspecting the model's generation function and examining the logits for all vocabulary items could clarify this behavior, such an analysis was not completed in this study. Therefore, the early dominance of *the* should be interpreted cautiously. Finally, greedy decoding may amplify the model's early preference for the definite determiner, and future work should examine whether alternative decoding strategies yield different developmental patterns in the model.

Acknowledgments

This research was supported by a University of Maryland Baggett Fellowship and by the University of Maryland Strategic Partnership: MPowering the State, a formal collaboration between the University of Maryland College Park and the University of Maryland Baltimore. We thank the computational cognitive science group for helpful comments and

discussion.

References

Sudha Arunachalam and Sandra R. Waxman. 2010. Language and conceptual development. *WIREs Cognitive Science*, 1(4):548–558.

Stella Biderman, Hailey Schoelkopf, Quentin Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, Aviya Skowron, Lintang Sutawika, and Oskar van der Wal. 2023. Pythia: A suite for analyzing large language models across training and scaling.

Amy E. Booth, Sandra R. Waxman, and Yi Ting huange. 2005. Conceptual information permeates word learning in infancy. *Developmental Psychology*.

Susan R. Braunwald. 1997. The development of because and so: Connecting language, thought and social understanding. In *Processing Interclausal Relationships in the Production and Comprehension of Text*. Lawrence Erlbaum Associate, Hillsdale, New Jersey.

Youjung Choi, Hyun joo Song, and Yuyan Luo. 2018. Infants' understanding of the definite/indefinite article in a third-party communicative situation. *Cognition*, 175:69–76.

Mark Davies. 2008. The corpus of contemporary american english.

Katherine Demuth, Jennifer Culbertson, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*.

Nathalie Fernandez, Rose Griffin, Patrick Shafto, and Naomi Feldman. 2024. Characterizing language learning trajectories with optimal transport. In *Paper presented at the Boston University Conference on Language Development*.

Zara Harmon and Vsevolod Kapatsinski. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, 98:22–44.

MP Maratsos. 1976. *The use of definite and indefinite reference in young children*. Cambridge University Press.

Rebecca Marvin and Tal Linzen. 2018. Targeted syntactic evaluation of language models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium. Association for Computational Linguistics.

Isabel Papadimitriou, Richard Futrell, and Kyle Mahowald. 2022. When classifying grammatical role, bert doesn't care about word order... except when it matters.

Michael Tomasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Alex Warstadt, Aaron Mueller, Leshem Choshen, Ethan Wilcox, Chengxu Zhuang, Juan Ciro, Rafael Mosquera, Bhargavi Paranjabe, Adina Williams, Tal Linzen, and Ryan Cotterell. 2023. Findings of the babylm challenge: Sample-efficient pretraining on developmentally plausible corpora. In *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, pages 1–34, Singapore. Association for Computational Linguistics.

Ken Wexler. 2011. Cues don't explain learning: Maximal trouble in the determiner system. *The processing and acquisition of reference*, 15.

Michael Wilson, Jackson Petty, and Robert Frank. 2023. How abstract is linguistic generalization in large language models? experiments with argument structure. *Transactions of the Association for Computational Linguistics*, 11:1377–1395.

Yuanfan Ying, Valentine Hacquard, Alexander Williams, and Jeffrey Lidz. 2024. Children do not overuse "the" in natural production. *Proceedings of the 48th annual Boston University Conference on Language Development*.