A Computational Simulation of Language Production in First Language Acquisition

Yuan Gao[1,2], Weiwei Sun[1],

{yg386,ws390}@cam.ac.uk

[1]Department of Computer Science & Technology, University of Cambridge
[2]ALTA Institute

Abstract

We introduce a computational framework for modeling child language production, focusing on the acquisition of the competence to map meaning onto linguistic form. Our approach uses graphs to formalize meaning and Synchronous Hyperedge Replacement Grammar (SHRG) to formalize the syntax–semantics interface. The setup provides computationally-sound induction algorithms of statistical grammar knowledge. We induce SHRGs solely from semantic graphs, and the resulting interpretable grammars are evaluated by their ability to generate utterances—providing a novel controlled paradigm to simulate child language acquisition. A notable finding is that unsupervised statistical learning (analogous to children's implicit learning mechanisms) performs as well as the corresponding supervised oracle when a proper symbolic grammar is assumed (reflecting knowledge gained via comprehension). The full dataset and code are available at https://github.com/yuan-w-gao/fla_hrg_induction.

1 Introduction

Children rapidly acquire the ability to produce context-appropriate utterances from sparse input (Chomsky, 1980; Berwick et al., 2011). This puzzling phenomenon calls for a scientific explanation, and computational models have drawn growing attention in this line of research. Existing computational studies of acquisition have largely prioritized isolated mechanisms, such as spreading activation in retrieval (Dell, 1986), sentence planning (Levelt et al., 1999), or production efficiency (Jaeger et al., 2012), over a more integrated goal: develop a computational system that can simulate the process by which children learn to produce. Production in acquisition still lacks a precise, testable account of how intended meanings are transformed into utterances, and how this ability is obtained.

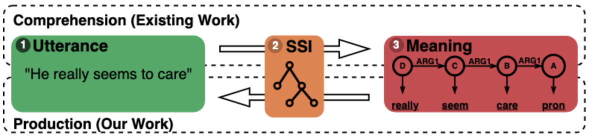


Figure 1: A conceptual bidirectional framework linking (1) *Utterance*, (2) *SSI* (syntax–semantics interface), and (3) *Meaning*. In comprehension (top), an utterance is interpreted into a structured meaning via the SSI; in production (bottom), a speaker starts from an intended meaning and constructs a sentence using the same interface. The dashed frame separates prior comprehension work from our production focus; both rely on the same interface.

This paper initiates a computational investigation into the developmental process of child language production. We formalize a new research task: given a structured meaning representation as input, learn or induce statistical grammatical knowledge that maps and predicts potential utterances. Our inquiry presupposes a bidirectional syntax–semantic interface (SSI)[1]. Figure 1 illustrates a simplified view of language comprehension and production with the three objects, utterances, the SSI, and meanings. In comprehension, the child interprets an observed utterance by deriving a structured meaning representation via hierarchical syntax. In production, this process is reversed: starting from an intended meaning, the child generates an

[1]By SSI we mean a structured mapping that connects a sentence's hierarchical derivation (syntax) to a compositional meaning representation (semantics). In NLP terms, the mapping is often operationalized in semantic parsing (form $\rightarrow$ meaning) and in surface realization/generation (meaning$\rightarrow$form). For example, in Combinatory Categorial Grammar, each lexical entry pairs a syntactic category with a typed semantic term; when categories combine, their meanings compose (typically by function application) to yield a logical form to encode truth-conditional semantics. The same syntax–semantics pairing information can be used in reverse for realization (Steedman, 2001; Zettlemoyer and Collins, 2005; Reiter, 2000). We use SSI as a neutral label for mapping between natural language utterances and compositional semantics through syntax, independent of grammar formalism.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 10992–11006
November 4-9, 2025 ©2025 Association for Computational Linguistics

Work	Dir.	Input	Interface (2)	MR (3)
Abend et al. (2017); Zettlemoyer and Collins (2005); Kwiatkowksi et al. (2010); Bisk and Hockenmaier (2012)	1 +3 → 2	Utterance + Meaning	CCG	Logical form
Beekhuizen et al. (2014)	1 +3 → 2	Utterance + Meaning	Construction Grammar	Graph-based
Chang (2008)	1 +3 → 2	Utterance + Meaning	Embodied Construction Grammar	Schema-based
Nevens et al. (2022); Gerasymova and Spranger (2010)	1 +3 → 2	Utterance + interaction	Fluid Construction Grammar	Incremental Recruitment Language
This work	3 → 2 /1	Meaning	SHRG	Graph-based

Table 1: Representative approaches to learning the mapping among the three objects in Figure 1. The Dir. column uses the indices (1=utterance, 2=SSI, 3=meaning). Interface names the SSI formalism; MR the meaning representation.

utterance through the same syntax—semantics interface.

This perspective poses a fundamental question for modeling production, how do we ground an inherently cognitive and abstract process to a computational task? More specifically, two main challenges are highlighted: 1) How to represent the speaker's mental intended meaning, and 2) How to learn the interface that maps this representation to form. To address these challenges, we follow three working hypotheses. **H1: Linguistic Relativism.** We adopt the Sapir-Whorf hypothesis (Whorf, 1956; Sapir, 1929) and assume that the speaker's mental representation can be fully captured (the strong version) or approximated (the weak version) by semantics codified by linguistic signals. **H2: Comprehension precedes production.** We assume that language production does not start from scratch, but builds upon grammatical knowledge already acquired through comprehension. This developmental path is widely accepted in psycholinguistics (Clark and Hecht, 1983; Bornstein and Hendricks, 2011; Benedict, 1979; Bauer et al., 2002; Lowenthal, 2012; Kern, 2007). **H3: Children are statistical learners.** Children are sensitive to frequency patterns in language use and acquire probabilistic mappings between meaning and form by tracking patterns in their linguistic environment (Waxman and Markow, 1995; Saffran et al., 1996; Erickson and Thiessen, 2015).

On the operational side, we formalize the syntax–semantics interface using Synchronous Hyperedge Replacement Grammar (SHRG; Drewes et al. 1997), following its success in SemBanking (Li et al., 2025), semantic parsing (Chen et al., 2018, 2019; Zhao et al., 2020) and natural language generation (Ye and Sun, 2020). The meaning representation is English Resource Semantics–style semantic graphs (Flickinger et al., 2014). Specially, we utilise Elementary Dependency Graph (EDS; Oepen et al., 2002). The concrete computational task is to induce probabilities of predefined SHRG rules from solely EDS graphs. Table 1 compares this task to previous work.

We conduct experiments on the child speech, rather than child-directed speech, from Child Language Data Exchange System (CHILDES; Macwhinney, 2000). English Resource Grammar (ERG; Flickinger, 2000) is applied for linguistically-precise syntactico-semantic parsing. We collect a dataset of 966,030 parsed utterances from 79 corpora (77.6% parsing coverage of 1,244,745 total utterances), spanning from birth to age 10. Experiments on this dataset demonstrate that statistical learning approaches effectively model language production in acquisition. Our Expectation Maximization (EM; Dempster et al., 1977) grammar induction achieves parsing accuracy (0.952) comparable to an oracle with perfect grammatical knowledge (0.948), with similar trends in generation quality (0.777 vs 0.688 in BLEU score).

Remarkably, the strong performance of the unsupervised model shows that, given a basic symbolic grammar, statistical learning through maximizing likelihood of meaning representations alone can approach the performance of a statistically optimal supervised model. This finding remains robust across different EM variants, levels of grammar abstraction, and even in an incremental one-pass learning scenario that more closely mimics naturalistic acquisition. These results also demonstrate the usefulness of graph-based meaning representation and graph grammars in studying child language acquisition.

2 Related Work

2.1 Grammar Induction

Grammar induction learns structured linguistic representations from raw data. Early work in NLP focused on probabilistic context-free grammar (PCFG) induction, using the Expectation-Maximization (EM) algorithm (Lari and Young, 1990; Carroll and Charniak, 1992; Pereira and Schabes, 1992). Bayesian approaches (Johnson et al., 2007; Cohn and Blunsom, 2010) introduced nonparametric priors, improving hierarchical structure inference. Neuralized grammar induction approaches learn latent hierarchical structures via end-to-end optimization (Dyer et al., 2016; Jin et al., 2018; Kim et al., 2019; Yang et al., 2021). Beyond PCFGs, distributional learning algorithms, inspired by formal language theory (Clark, 2010), infer grammars based on substring statistics and structural regularities. Minimum Description Length methods (de Marcken, 1996) frame induction as a compression problem.

2.2 Modeling of Language Acquisition

Construction grammar and CCG-based models represent a significant branch of computational learning of grammars. Table 1 categorizes some notable prior work based on the learning task and formalisms. It is worth to point out that most existing work require the utterance as input (Abend et al., 2017; Zettlemoyer and Collins, 2005; Kwiatkowksi et al., 2010; Bisk and Hockenmaier, 2012; Beekhuizen et al., 2014; Chang, 2008; Nevens et al., 2022; Gerasymova and Spranger, 2010; Chen and Mooney, 2008; Dominey and Boucher, 2005; Doumen et al., 2023; Dunn, 2017), making them less cognitively plausible for language production simulations where the utterance could not have been known prior to production.

3 Representing Meaning of Child Speech

3.1 Graphs as Meaning Representations

Following H1 and building on established linguistic theories of the syntax–semantics interface, we treat meaning representations derived from child speech as the input to our production model. In our framework, we adopt graph-structured semantic representations, following a long tradition in computational semantics that views linguistic meaning as compositional, typed, and relational. Graph-based representations offer a flexible and expressive format for encoding predicate–argument structure, scope, and modification, and have been widely used in frameworks such as Abstract Meaning Representation (AMR; Banarescu et al., 2013) and English Resource Semantics (ERS; Flickinger et al., 2014).

In this work, we use ERS, the semantic analyses produced by English Resource Grammar (ERG; Flickinger, 2000), a linguistically precise, broad-coverage grammar of English. In particular, we choose the format of Elementary Dependency Structures (EDS; Oepen et al., 2002), which represents meanings by labeled graphs. See an example in Figure 2. Semantic parsing with ERG into EDS achieved 0.973 overall parsing F1 score on various benchmark datasets, due to their highly specific and precise nature (Oepen and Flickinger, 2019; Oepen et al., 2020).

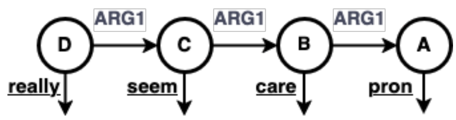


Figure 2: Example ERG-derived meaning representation for the utterance *He really seems to care*. Single node hyperedges represent concepts, while normal edges represent predicate–argument relationships.

While the ERG was developed for well-formed adult English, we argue that it is a reasonable choice for modeling of child language. First, ERG achieves high parsing coverage for a broad range of child utterances (see §3.2), particularly from the telegraphic stage onward. Parsing failures in earlier stages often result from disfluencies, fragmentary input, or nonstandard forms rather than true structural mismatch. It is reasonable to exclude such data from training. Second, ERG offers a principled type hierarchy and interpretable derivations that support abstraction: fine-grained rule and type distinctions can be systematically projected onto coarser categories to simulate early-stage grammars. Experiments suggest the model and the representation remain robust against different level of grammar abstractions (see §6.4).

3.2 Dataset Construction

We construct a from CHILDES (Macwhinney, 2000). Our dataset pairs transcribed child utterances with semantic graphs obtained via ERG parsing. The semantic graphs are utilized as the input for our simulation experiments. In total, 1,244,745 utterances were selected from 18 UK and 61 North American corpora covering ages between 0 and 10 to capture key language acquisition stages[2]. 966,030 utterances were successfully parsed with 77.6% parsing coverage. Figure 3 shows parsing

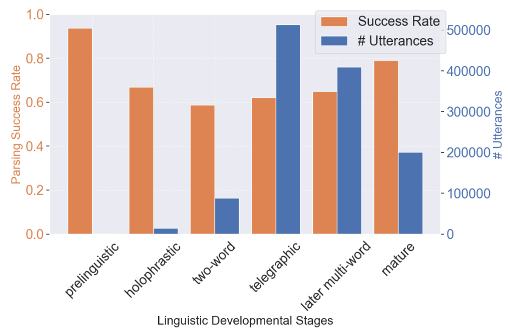


Figure 3: Parsing success rate and total number of utterances by the 7 language acquisition stages: the prelinguistic stage (0–12 months), the holophrastic stage (12–18 months), the two-word stage (18–24 months), the telegraphic stage (24–36 months), the later multi-word stage (36–60 months), and the mature stage (60+ months) (Ingram, 1989; de Villiers and de Villiers, 1979; Rudman and Titjen, 2018).

success rates across language acquisition stages. Success is the highest in the prelinguistic stage (0–12 months), possibly due to the simplicity of vocalizations and the small amount of data (464 utterances in total). Success rate declines in the holophrastic and two-word stages, likely due to non-standard or ungrammatical constructions and novel word forms. From the telegraphic stage onward, success improves as children's syntax becomes more structured.

[2]Full list of corpora used are in Appendix A

4 The Grammar Formalism

To mediate between meaning and syntax, we use SHRG, a synchronous version of Hyperedge Replacement Grammar (HRG; Drewes et al., 1997), to synchronize semantic and syntactic derivations. Figure 4 demonstrates the syntactico-semantic derivation for the running example.

Compositional Semantics via SHRG. In an SHRG derivation tree, each node corresponds to a syntactic constituent and is paired with a local semantic subgraph representing that constituent's meaning. The semantic composition proceeds along the syntactic hierarchy: whenever two child constituents combine under a parent in the parse tree, their associated meaning graphs are glued together by the corresponding grammar rule to form a larger graph. The smallest semantic units are word-level graphs provided by lexical rules in an HRG. In Figure 4, γ_4, γ_6, γ_7 and γ_8. These lexical rules/graphs encode predicate—argument structures of the paired content words — much like lexical entries in Combinatory Categorial Grammar (CCG; Steedman, 1996, 2000) carries its own logical formula. Figure 5 shows a comparison of a transitive verb 'like'.

Phrasal rules in SHRG (e.g. combining a noun phrase and a verb phrase into a sentence) perform the "glue" operations: they identify the shared nodes between child semantic graphs (such as merging the predicate's slot with the subject's node, i.e., γ_1 in Figure 4) and thereby yield a unified graph for the parent constituent. These glue operations are explicitly encoded in the HRG production rules as the connection of external nodes in the graphs. In short, each syntactic combination corresponds to a semantic combination. This design ensures that the meaning of the whole utterance is built compositionally from the meanings of its parts.

Probabilistic SHRG A probabilistic SHRG turns the symbolic grammar into a generative model: every rule in the grammar gets a weight that reflects how likely it is to be used. The probability of a full derivation is just the product of the rule weights along the way, like in a probabilistic context-free grammar (PCFG). This means the model can learn preferences over how things are said. Probabilistic SHRG gives us a way to represent not just what's possible in the grammar, but also what's likely, which is exactly the kind of sen-

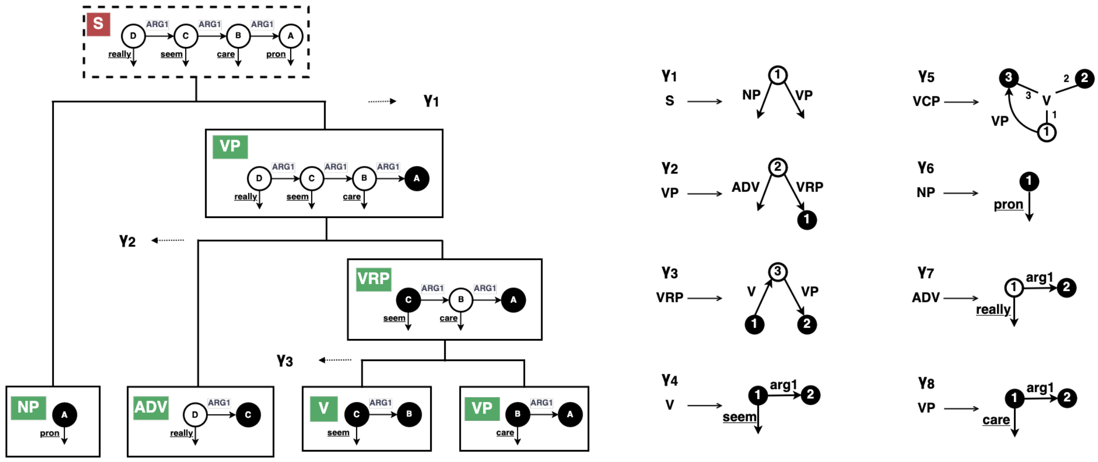


Figure 4: An HRG-based syntactico–semantic derivation for *He really seems to care.* The semantic graph (top left) is recursively decomposed into meaning constituents using HRG rules (right), while each subgraph is aligned to a syntactic category (green/red labels). External nodes (filled black nodes) mark nodes that are 'available' for subgraph merging; edge labels and node indices maintain semantic and syntactic alignment.

$(\mathsf{S}\backslash\mathsf{NP_y})/\mathsf{NP_x}$

$\lambda x.\lambda y.\underline{\text{like}}(y, x)$

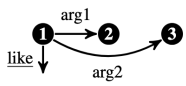


Figure 5: Example HRG rule and its corresponding CCG rule. The external nodes 1, 2, and 3 in the HRG rule correspond to the syntactic categories S, NP_y, and VP_x.

sitivity we want if we're modeling how kids learn to speak from examples in their environment.

Probabilistic SHRG is compatible with inference and learning algorithms originally developed for PCFGs. In particular, the inside–outside algorithm—used to efficiently compute the expected counts of grammar rules—extends naturally to the graph setting, allowing us to calculate the likelihood of all possible derivations for a meaning graph under the current grammar. Combined with Expectation–Maximization (EM), this enables unsupervised learning: the model updates rule probabilities to better match the observed data, even without gold derivations. This makes probabilistic SHRG well-suited for simulating language acquisition.

Algorithmic Advantages An important advantage of SHRG is that it is a generative-enumerative grammar formalism (Pullum and Scholz, 2001), enabling algorithmically-efficient conversion in both directions between meanings and surface strings. In principle, dynamic programming can be applied to parse a semantic graph into all possible derivation trees (and hence generate all compatible utterances), or conversely to parse a sentence into all possible semantic graphs (Chiang et al., 2013; Ye and Sun, 2020). In particular, low-degree polynomial parsing algorithms are available. Therefore, SHRG is suitable in computational practice as a bidirectional interface: the same grammar that maps an input meaning to a well-formed sentence can also parse an input sentence to recover its semantic representation.

5 Unsupervised Learning Algorithms

Using the probabilistic SHRG formalism, learning reduces to estimating rule probabilities from semantic graphs. However, since we cannot observe the derivation trees themselves, we treat this as an unsupervised grammar induction problem and apply EM to maximize data likelihood.

5.1 Expectation-Maximization (EM)

Grammar induction is formulated as unsupervised learning over grammar parameters ϕ. Given observed semantic graphs $G = \{g_1, ..., g_n\}$, we maximize the likelihood function:

$$p(G; \phi) = \prod_{i=1}^{n} p(g_i; \phi) \tag{1}$$

where the joint probability of a graph g and its parse tree t factorizes as:

$$p(g, t; \phi) = \prod_{\gamma \in P} \phi(\gamma)^{c(\gamma; t, g)} \tag{2}$$

where $c(\gamma; t, g)$ is the count of rule γ in parse t.

Computing Expectation The E-step computes expected rule counts:

$$\mathbb{E}_{\phi}[c(\gamma;t,g)] = \sum_{t} c(\gamma;t,g)p(t|g;\phi) \quad (3)$$

where $p(t|g,\phi) = p(g,t|\phi)/p(g|\phi)$. We compute this efficiently using inside-outside probabilities: The expected count for rule γ is then:

$$\mathbb{E}_{\phi}[c(\gamma;t,g)] = \frac{\phi(\gamma)}{p(g;\phi)} \sum_{\beta\in\beta_g(\gamma)} \delta(\beta) \prod_{s_i\in c(\beta)} \alpha(s_i) \quad (4)$$

where α and δ are inside and outside probabilities, respectively.

Parameter Updates The M-step applies maximum likelihood estimation:

$$\phi^*(\gamma = A \to r) = \frac{\mathbb{E}_{\phi}[c(\gamma)]}{\sum_{\gamma'\in P_A} \mathbb{E}_{\phi}[c(\gamma')]} \quad (5)$$

5.2 Viterbi EM

Unlike standard EM, which marginalizes over all parses, Viterbi EM selects only the most probable parse t^*:

$$t_i^* = \arg\max_{t} p_{\phi}(g_i, t) \quad (6)$$

5.3 Batch EM

Batch EM improves efficiency by updating parameters more frequently. Given a dataset, we partition it into B of size $m = n/B$. For each batch $\mathcal{G}_i$, we compute local expected counts.

The parameter update is a weighted combination:

$$\phi^*(\gamma) = \alpha\phi^i(\gamma) + (1-\alpha)\phi^{i-1}(\gamma) \quad (7)$$

where $\alpha = m/n$ controls batch weighting. A final normalization step ensures a proper probability distribution:

$$\phi(\gamma = A \to r) = \frac{\exp(\phi(\gamma))}{\sum_{\gamma'\in P_A} \exp(\phi^*(\gamma'))} \quad (8)$$

5.4 Online EM

Online EM (Liang and Klein, 2009) updates parameters after processing each graph g_t. The learning rate decreases with time:

$$\alpha_t = \frac{1}{\sqrt{t}} \quad (9)$$

For each g_t, we compute expected counts and apply an online update:

$$\phi^t(\gamma) = \alpha_t\phi^{new}(\gamma) + (1-\alpha_t)\phi^{t-1}(\gamma) \quad (10)$$

where $\phi^{new}(\gamma)$ is computed from the current example. Normalization is applied after each update:

$$\phi^t(\gamma = A \to r) = \frac{\exp(\phi^t(\gamma))}{\sum_{\gamma'\in P_A} \exp(\phi^t(\gamma'))} \quad (11)$$

6 Experimental Results and Analysis

These experiments examine grammar acquisition across developmental stages using SHRG grammar induction. Spanning birth (0 months) to 120 months (10 years), our evaluation compares model performance against an oracle and a baseline. The oracle is a supervised approach that estimates rule probabilities from gold-standard parse trees and the baseline uniformly samples valid rules at each derivation node.

We model production at the population level due to both data sparsity and theoretical motivation. Most children in CHILDES contribute too few utterances for reliable individual modeling, and child language development follows consistent trends that generalize across individuals and languages (Brown, 1973). Our approach captures these broad patterns while remaining extensible to individual learners given sufficient longitudinal data.

6.1 Evaluation

We evaluate our grammar induction using parsing accuracy and generation quality, offering complementary perspectives on model performance.

Parsing accuracy measures how well the induced grammar identifies meaning constituents in semantic graphs. Specifically, it compares the subgraphs (meaning constituents) found by the induced derivations with those in the gold-standard parses. The metric is computed as the exact match rate, and we report the F1 score. **Bilingual Evaluation Understudy** (BLEU; Papineni et al., 2002) is computed against original child utterances to evaluate how well the induced grammar can produce utterances from a given meaning.

6.2 Oracle Performance

Before presenting our main experimental results, we first examine the performance of our oracle—a fully supervised model that estimates rule probabilities from gold-standard parse trees. One concern of this work is the dataset derived from CHILDES, which pairs child utterances with automatically generated meaning representations with a reliance on adult grammar. As shown in Table 2, parsing F1

Dev. Stage	Parsing F1			BLEU		
	EM	Oracle	Baseline	EM	Oracle	Baseline
Prelinguistic	0.996	0.996	0.506	0.955	0.955	0.952
Holophrastic	0.990	0.990	0.592	0.916	0.916	0.907
Two-word	0.987	0.987	0.574	0.882	0.881	0.824
Telegraphic	0.981	0.980	0.554	0.870	0.868	0.806
Later Multi-word	0.979	0.972	0.467	0.818	0.813	0.687
Mature	0.982	0.976	0.372	0.846	0.846	0.714

Table 2: Parsing F1 and BLEU scores across language acquisition stages.

scores for the oracle range from 0.996 for the prelinguistic stage to 0.976 for the mature stage, with correspondingly strong BLEU scores for generation.

This strong oracle performance validates two key aspects of our framework. First, it confirms that ERG-derived semantic hypergraphs provide an effective representation of meaning for modeling child language production. Second, it demonstrates that SHRG offers a powerful formalism for capturing the syntax–semantics interface in developing grammars.

6.3 Main Results

Table 2 reports the parsing F1 and BLEU scores across different developmental stages. The EM model achieves near-identical parsing F1 scores to the oracle across all stages. The baseline, which has access to the same assumed grammatical knowledge but uses uniform rule probabilities (without statistical learning), performs substantially worse. This suggests that maximizing the likelihood of all potential analyses is effective at learning structured linguistic patterns from child speech, achieving robust generalization despite the lack of explicit supervision. It also demonstrates that the learning component is crucial and the task is not trivially solved by the provided symbolic grammatical knowledge alone.

An important observation is that the parsing F1 trends align with developmental patterns in language acquisition. While performance remains high throughout, a slight dip occurs during key transitional stages—holophrastic, two-word, telegraphic, and later multi-word stages—before recovering in the mature stage. This pattern is consistent with the increased variability and complexity in early multi-word speech, which poses greater challenges for parsing (Valian, 2024).

BLEU scores follow a similar trend. This suggests that the EM model not only learns grammatically sound structures but also captures frequent surface patterns in child speech, leading to more naturalistic utterance generation.

6.4 Grammar Granularity

The strong performance of the EM algorithm raises the question of whether the fine-grained grammar structure makes the task artificially easy. To test this, we introduce four levels of rule granularity, progressively reducing syntactic specificity.

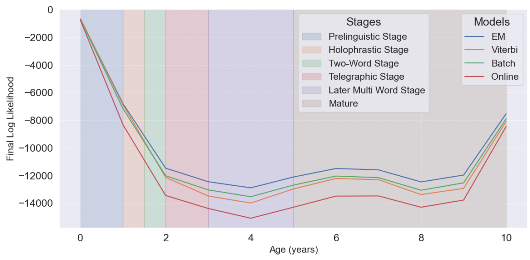
Grammar	Parsing F1		BLEU	
	EM	Or.	EM	Or.
Fine	0.952	0.948	0.777	0.688
Medium Fine	0.940	0.942	0.734	0.626
Medium Coarse	0.938	0.934	0.725	0.607
Coarse	0.920	0.906	0.651	0.598

Table 3: Parsing F1 and BLEU scores across different grammar granularities. Results are averaged across all age groups for clarity.

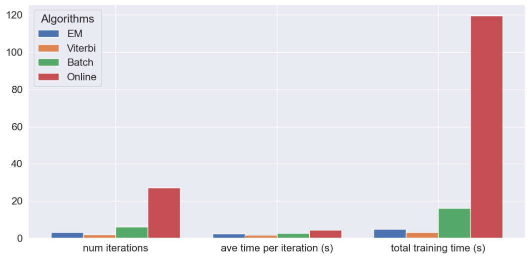
At the most detailed level, rules encode full subcategorization and head information (e.g., `VP/N@V`, indicating a verb phrase missing a noun, headed by a verb). We then remove head annotations (`VP/N`), followed by subcategorization (`V`), and finally collapse all categories into three coarse types: `VP`, `NP`, and `AP`.

Table 3 shows that while parsing accuracy drops with reduced granularity, the model remains close to oracle performance across all settings. Generation performance remains strong, with BLEU scores at the coarsest level comparable to the oracle with the finest-grained grammar. These results demonstrate that our model generalizes well even under minimal syntactic structure, confirming the robustness of the findings in Table 2.

6.5 EM Varieties



(a) Final log likelihood of the training corpus



(b) Training Time

Figure 6: Training statistics of different EM algorithms.

Figure 6a shows final log-likelihood scores for different EM variants across age groups. Standard EM performs best, followed by Batch EM and Online EM. This ordering reflects their update strategies: standard EM uses full marginalization, Batch EM smooths over mini-batches, and Online EM updates incrementally, making it more prone to local optima. The spike at age 10 is likely due to data sparsity, as only one monthly datapoint exists in that group.

Algorithm	Parsing F1		BLEU	
	EM	Or.	EM	Or.
EM	0.952	0.948	0.777	0.688
Viterbi	0.945	0.948	0.773	0.688
Batch	0.952	0.948	0.776	0.688
Online	0.952	0.948	0.776	0.688

Table 4: Parsing accuracy and BLEU scores of different EM algorithms. Results are averaged across all age groups for clarity.

Table 4 compares the performance of different EM algorithm variants. All variants achieve remarkably similar performance. The similar performance suggests that the choice of update strategy - whether using full marginalization (standard EM), batched updates, or online learning - has minimal impact on the model's ability to identify correct grammatical structures.

6.6 One-Pass Online Learning

While previous experiments rely on multiple passes over the data, human learners acquire language incrementally—processing each utterance once without repeated exposure to the same sentences in a controlled manner. To simulate this, we evaluate the model in a more realistic single-pass setting using the online EM algorithm.

In this setup, the model updates its parameters after each utterance and sees the data only once, mimicking naturalistic acquisition. As shown in Figure 7, performance declines slightly compared to the oracle, but remains strong overall. This demonstrates that the model can learn accurate structure from minimal exposure, highlighting its capacity to generalize in realistic learning conditions.

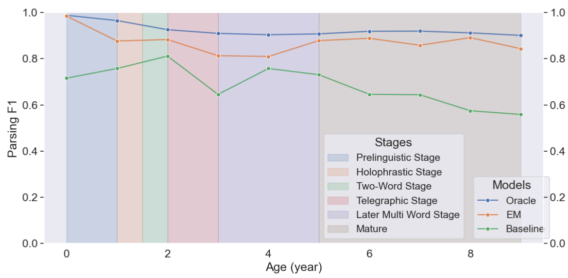


Figure 7: Parsing F1 of the Online EM algorithm after 1 iteration

7 Discussion

7.1 What the simulations establish, and what they do not

Our results from the previous section show that, given a symbolic interface between meaning and form, estimating rule weights from meaning representations alone yields a probabilistic grammar that performs well on production tasks within the same interface. Across developmental stages, EM reaches parsing F1 essentially identical to the supervised oracle, and achieves strong generation score; the uniform baseline, exposed to the same symbolic rules but without statistical learning, lags well behind. This pattern holds across grammar granularities, EM variants, and even in a one-pass setting, indicating that the learning signal in meaning representations is sufficient to tune a grammar for production under our assumptions. This shows that if children had access to such an interface from comprehension, then attempts to produce could, in principle, supply enough signal to tune production preferences.

What we have not shown is a learning mechanism in children. These simulations do not claim that children implement EM or that production-driven estimation alone explains language production. Thus, the contribution is not a claim about how children learn, but a computational check of sufficiency: under explicit and rich representational commitments, the production side can be tuned from meaning-only supervision.

7.2 What computational simulation contributes to acquisition research

Even though the simulations do not identify the exact learning mechanism in children, they still offer insights for acquisition research by showing what follows from specific representational assumptions. First, they turn a qualitative hypothesis into a quantitative test of sufficiency: under a fixed interface, meaning-only estimation yields a grammar that support production, showing that annotated derivations are not required in principle and shifting the burden of explanation onto what the interface must provide. Second, because the interface is modular, the same setup lets of locate the necessary properties of that interface: when we vary rule granularity or lexicalization and hold the evaluation constant, the points at which performance remains stable versus collapses identify which representational commitments actually do the explanatory work, rather than being artifacts of training. Third, a single parameterization serves both parsing and generation, which yields linked, falsifiable predictions. In this sense, the simulations are not a model of how children learn; they are a conservative test of what would be enough, one that surfaces the assumptions that matter and turns them into empirical targets for future data and experiment.

7.3 Toward Bilingualism and Second Language Acquisition

Bilingualism and second language acquisition (SLA) are central to acquisition research. They expand the acquisition theories by requiring accounts of learning with prior linguistic knowledge and potentially multiple interacting grammars, refining the inductive bias, and learner plasticity. Because these phenomena set boundary conditions on learnability and change, a general theory should treat bilingual and L2 learning as core tests.

Extending our L1-focused setup to bilingualism/L2, however, is non-trivial. First, high-precision, broad-coverage grammars and reliable parsers for non-English languages are scarce, limiting cross-lingual coverage. Second, while our experimental design is a necessary step toward modeling SLA, it is not directly translatable: additional factors must be incorporated, including L1 grammar transfer, L1/L2 interference, and interlanguage grammar. Third, prior work explicitly coupling an interpretable grammar with production-driven estimation for SLA is limited. We do not address these topics in this paper.

Even so, the present framework provides the right primitives and tools for future SLA work. By pairing an explicit, precise grammar with production-driven estimation, it yields interpretable, construction-level grammar profiles that can be compared across languages and tracked over learner proficiency. This level of interpretability enables falsifiable, theory-grounded analyses of how individual grammars shift at the syntax–semantics interface, and it supports principled connections to education and assessment.

8 Conclusion

In this work, our model demonstrates strong learning and generation capabilities, achieving parsing F1 scores comparable to an oracle and producing high-quality utterances. Second, these results hold across different grammar types, showing that the model robustly adapts to variations in syntactic representation. However, the model remains constrained by the structure of the given grammar and does not independently refine or generalize beyond its predefined rules. Finally, we show that this framework enables controlled simulations of language acquisition, providing a computational tool for studying developmental patterns.

9 Limitations

One limitation of this work is the dataset derived from CHILDES, which pairs child utterances with automatically generated meaning representations. While this dataset enables large-scale computational modeling of language production, its reliance on an adult parser affects coverage of child-specific constructions, and data distribution remains uneven across age groups. Additionally, the focus on English limits cross-linguistic applicability, and the absence of prosodic and gestural information restricts the multimodal aspects of language acquisition. Addressing these challenges in future work will further enhance the realism and generalizability of computational models of child language learning. Finally, the lack of longitudinal data of individual children restricted the experiments to macro-level coarse-grained language production simulation at the population level. However, our computational framework remains general and flexible. With sufficient longitudinal data, it could be adapted to model individual learners with minimal changes.

10 Acknowledgements

This paper reports on research partly supported by Cambridge University Press & Assessment.

We thank the anonymous reviewers and the area chair for their useful feedback and suggestions, which greatly improved the manuscript. We thank Wenxi Li, Andrew Caines, and Nils Lauermann for reviewing earlier drafts of this paper.

References

Omri Abend, Tom Kwiatkowski, Nathaniel J. Smith, Sharon Goldwater, and Mark Steedman. 2017. Bootstrapping language acquisition. Cognition, 164:116–143.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Elizabeth Bates, Inge Bretherton, and Lynn Sebestyen Snyder. 1991. From First Words to Grammar: Individual Differences and Dissociable Mechanisms. Cambridge University Press.

Daniel J. Bauer, Beverly A. Goldfield, and J. Steven Reznick. 2002. Alternative approaches to analyzing individual differences in the rate of early vocabulary development. Applied Psycholinguistics, 23(3):313–336.

Barend Beekhuizen, Rens Bod, Afsaneh Fazly, Suzanne Stevenson, and Arie Verhagen. 2014. A Usage-Based Model of Early Grammatical Development. In Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics, pages 46–54, Baltimore, Maryland, USA. Association for Computational Linguistics.

David C. Bellinger and Jean B. Gleason. 1982. Sex differences in parental directives to young children. Sex Roles: A Journal of Research, 8(11):1123–1139.

Helen Benedict. 1979. Early lexical development: Comprehension and production. Journal of Child Language, 6(2):183–200.

M. M. Berl, L. M. Balsamo, B. Xu, E. N. Moore, S. L. Weinstein, J. A. Conry, P. L. Pearl, B. C. Sachs, C. B. Grandin, C. Frattali, F. J. Ritter, S. Sato, W. H. Theodore, and W. D. Gaillard. 2005. Seizure focus affects regional language networks assessed by fMRI. Neurology, 65(10):1604–1611.

Robert C. Berwick, Paul Pietroski, Beracah Yankama, and Noam Chomsky. 2011. Poverty of the Stimulus Revisited. Cognitive Science, 35(7):1207–1242.

Yonatan Bisk and Julia Hockenmaier. 2012. Simple robust grammar induction with combinatory categorial grammars. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, AAAI'12, pages 1643–1649, Toronto, Ontario, Canada. AAAI Press.

Lynn S. Bliss. 1988. Modal usage by preschool children. Journal of Applied Developmental Psychology, 9(3):253–261.

Lois Bloom. 1970. Language development: Form and function in emerging grammars.

John N. Bohannon and Angela L. Marquis. 1977. Children's control of adult speech. Child Development, 48(3):1002–1008.

Marc H. Bornstein and Charlene Hendricks. 2011. Basic language comprehension and production in >100,000 young children from sixteen developing nations. Journal of child language, 39(4):899.

Susan R. Braunwald. 1971. Mother-Child Communication: The Function of Maternal-Language Input. WORD, 27(1-3):28–50.

M. R. Brent and J. M. Siskind. 2001. The role of exposure to isolated words in early vocabulary development. Cognition, 81(2):B33–44.

Roger Brown. 1973. Development of the first language in the human species. American Psychologist, 28(2):97–106.

Roger Brown. 2013. A First Language: The Early Stages. Harvard University Press.

Glenn Carroll and Eugene Charniak. 1992. Two experiments on learning probabilistic dependency grammars from corpora. Citeseer.

Nancy Chih-Lin Chang. 2008. Constructing Grammar: A Computational Model of the Emergence of Early Constructions. Ph.D. thesis, University of California, Berkeley.

David L. Chen and Raymond J. Mooney. 2008. Learning to sportscast: A test of grounded language acquisition. In Proceedings of the 25th International Conference on Machine Learning, ICML '08, pages 128–135, New York, NY, USA. Association for Computing Machinery.

Yufei Chen, Weiwei Sun, and Xiaojun Wan. 2018. Accurate SHRG-based semantic parsing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 408–418, Melbourne, Australia. Association for Computational Linguistics.

Yufei Chen, Yajie Ye, and Weiwei Sun. 2019. Peking at MRP 2019: Factorization- and composition-based parsing for elementary dependency structures. In Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning, pages 166–176, Hong Kong. Association for Computational Linguistics.

David Chiang, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, Bevan Jones, and Kevin Knight. 2013. Parsing graphs with hyperedge replacement grammars. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 924–932, Sofia, Bulgaria. Association for Computational Linguistics.

Noam A. Chomsky. 1980. Rules and Representations. Behavioral and Brain Sciences, 3(127):1–61.

Alexander Clark. 2010. Distributional Learning of Some Context-Free Languages with a Minimally Adequate Teacher. In Grammatical Inference: Theoretical Results and Applications, pages 24–37, Berlin, Heidelberg. Springer.

Eve Clark. 1978. Awareness of Language: Some Evidence from what Children Say and Do. pages 17–43.

Eve V. Clark and Barbara F. Hecht. 1983. Comprehension, production, and language acquisition. Annual Review of Psychology, 34:325–349.

Trevor Cohn and Phil Blunsom. 2010. Blocked Inference in Bayesian Tree Substitution Grammars. In Proceedings of the ACL 2010 Conference Short Papers, pages 225–230, Uppsala, Sweden. Association for Computational Linguistics.

Gina Conti-Ramsden and Jane Dykins. 1991. Mother-child interactions with language-impaired children and their siblings. International Journal of Language & Communication Disorders, 26(3):337–354.

Alan Cruttenden. 1978. Assimilation in child language and elsewhere. Journal of Child Language, 5(2):373–378.

Carl de Marcken. 1996. Linguistic Structure as Composition and Perturbation. In 34th Annual Meeting of the Association for Computational Linguistics, pages 335–341, Santa Cruz, California, USA. Association for Computational Linguistics.

Peter A. de Villiers and Jill G. de Villiers. 1979. Early Language. Harvard University Press.

Gary S. Dell. 1986. A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3):283–321.

M. J. Demetras, Kathryn Nolan Post, and Catherine E. Snow. 1986. Feedback to first language learners: The role of repetitions and clarification questions. Journal of Child Language, 13(2):275–292.

Marty Demetras. 1989. Working parents conversational responses to their two-year-old sons.

A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1):1–22.

David K. Dickinson and Patton O. Tabors, editors. 2001. Beginning Literacy with Language: Young Children Learning at Home and School. Beginning Literacy with Language: Young Children Learning at Home and School. Paul H. Brookes Publishing Co., Baltimore, MD, US.

Peter Ford Dominey and Jean-David Boucher. 2005. Learning to talk about events from narrated video in a construction grammar framework. Artificial Intelligence, 167(1):31–61.

Jonas Doumen, Katrien Beuls, and Paul Van Eecke. 2023. Modelling Language Acquisition through Syntactico-Semantic Pattern Finding. In Findings of the Association for Computational Linguistics: EACL 2023, pages 1347–1357, Dubrovnik, Croatia. Association for Computational Linguistics.

Frank Drewes, Hans-Joerg Kreowski, and Annegret Habel. 1997. Hyperedge replacement graph grammars. pages 95–162.

Jonathan Dunn. 2017. Computational learning of construction grammars. Language and Cognition, 9(2):254–292.

Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent Neural Network Grammars. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 199–209, San Diego, California. Association for Computational Linguistics.

Lucy C. Erickson and Erik D. Thiessen. 2015. Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. Developmental Review, 37:66–108.

Andrea Feldman and Lise Menn. 2003. Up close and personal: A case study of the development of three English fillers. Journal of child language, 30(4):735–768.

Paul Fletcher and Michael Garman. 1988. Normal language development and language impairment: Syntax and beyond. Clinical Linguistics & Phonetics, 2(2):97–113.

Dan Flickinger. 2000. On building a more effcient grammar by exploiting types. Natural Language Engineering, 6(1):15–28.

Dan Flickinger, Emily M. Bender, and Stephan Oepen. 2014. Towards an encyclopedia of compositional semantics: Documenting the interface of the English Resource Grammar. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 875–881, Reykjavik, Iceland. European Language Resources Association (ELRA).

Michael A. Forrester. 2002. Appropriating Cultural Conceptions of Childhood: Participation in Conversation. Childhood-A Global Journal of Child Research, 9(3):255–276.

Catherine Garvey and Robert Hogan. 1973. Social Speech and Social Interaction: Egocentrism Revisited. Child Development.

V. C. Gathercole. 1986. The acquisition of the present perfect: Explaining differences in the speech of Scottish and American children. Journal of Child Language, 13(3):537–560.

S. A. Gelman, J. D. Coley, K. S. Rosengren, E. Hartman, and A. Pappas. 1998. Beyond labeling: The role of maternal input in the acquisition of richly structured categories. Monographs of the Society for Research in Child Development, 63(1):I–V, 1–148; discussion 149–157.

Kateryna Gerasymova and Michael Spranger. 2010. Acquisition of Grammar in Autonomous Artificial Systems. In Proceedings of the 2010 Conference on ECAI 2010: 19th European Conference on Artificial Intelligence, pages 923–928, NLD. IOS Press.

Ronald Bradley Gillam and Nils A Pearson. 2004. Test of narrative language. Pro-ed Austin, TX.

Pamela A. Hadley, Matthew Rispoli, Janet K. Holt, Colleen Fitzgerald, and Alison Bahnsen. 2014. Growth of finiteness in the third year of life: Replication and predictive validity. Journal of speech, language, and hearing research: JSLHR, 57(3):887–900.

Laura C. G. Haggerty. 1930. What a Two-and-One-Half-Year-Old Child Said in One Day. The Pedagogical Seminary and Journal of Genetic Psychology.

William S Hall, William E Nagy, and Robert L Linn. 1984. Spoken words: Effects of situation and social group on oral word usage and frequency. (No Title).

A. Henry. 1995. Belfast English and standard English : Dialect variation and parameter setting.

Roy Patrick Higginson. 1985. Fixing: Assimilation in language acquisition. (No Title).

Van Houten and Lori J. 1986. The Role of Maternal Input in the Acquisition Process: The Communicative Strategies of Adolescent and Older Mothers with the Language Learning Children. Technical report.

C. Howe. 1981. Acquiring language in a conversational context.

David Ingram. 1989. First Language Acquisition: Method, Description and Explanation. Cambridge University Press.

T. Florian Jaeger, Katrina Furth, and Caitlin Hilliard. 2012. Incremental Phonological Encoding during Unscripted Sentence Production. Frontiers in Psychology, 3:481.

Lifeng Jin, Finale Doshi-Velez, Timothy Miller, William Schuler, and Lane Schwartz. 2018. Unsupervised Grammar Induction with Depth-bounded PCFG. Transactions of the Association for Computational Linguistics, 6:211–224.

Mark Johnson, Thomas Griffiths, and Sharon Goldwater. 2007. Bayesian Inference for PCFGs via Markov Chain Monte Carlo. In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pages 139–146, Rochester, New York. Association for Computational Linguistics.

L. Kelly, E. Nixon, and J. Quigley. 2020. Father-child shared book-reading in light- and heavy-text conditions.

Sophie Kern. 2007. Lexicon development in French-speaking infants. First Language, 27(3):227–250.

Yoon Kim, Chris Dyer, and Alexander Rush. 2019. Compound Probabilistic Context-Free Grammars for Grammar Induction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 2369–2385, Florence, Italy. Association for Computational Linguistics.

Stan A. Kuczaj. 1977. The acquisition of regular and irregular past tense forms. Journal of Verbal Learning & Verbal Behavior, 16(5):589–600.

Tom Kwiatkowksi, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 1223–1233, Cambridge, MA. Association for Computational Linguistics.

K. Lari and S. J. Young. 1990. The estimation of stochastic context-free grammars using the Inside-Outside algorithm. Computer Speech & Language, 4(1):35–56.

W. J. Levelt, A. Roelofs, and A. S. Meyer. 1999. A theory of lexical access in speech production. The Behavioral and Brain Sciences, 22(1):1–38; discussion 38–75.

Wenxi Li, Xihao Wang, and Weiwei Sun. 2025. Compositional syntactico-SemBanking for English as a second or foreign language. In Findings of the Association for Computational Linguistics: ACL 2025, pages 24395–24406, Vienna, Austria. Association for Computational Linguistics.

Percy Liang and Dan Klein. 2009. Online em for unsupervised models. In Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics, pages 611–619.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. 2009a. Two-year-old children's production of multiword utterances: A usage-based analysis. 20(3):481–507.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. 2009b. Two-year-old children's production of multiword utterances: A usage-based analysis. In Cognitive Linguistics, volume 20.

F. Lowenthal. 2012. Language and Language Acquisition. Springer Science & Business Media.

Brian Macwhinney. 2000. The CHILDES project: Tools for analyzing talk. Child Language Teaching and Therapy, 8.

Brian MacWhinney and Catherine Snow. 1990. The Child Language Data Exchange System: An update. Journal of Child Language, 17(2):457–472.

Lorraine McCune. 1995. A normative study of representational play in the transition to language. Developmental Psychology, 31(2):198–206.

Michelle McGillion, Julian M. Pine, Jane S. Herbert, and Danielle Matthews. 2017. A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 58(10):1122–1131.

Aliyah Morgenstern, Marion Blondel, Pauline Beaupoil, Sandra Benazzo-Hourdel, Dominique Boutet, Angelika Kochan, and Fanny Limousin. 2018. Chapter 16. The blossoming of negation in gesture, sign and oral productions. In Maya Hickmann, Edy Veneziano, and Harriet Jisa, editors, Sources of Variation in First Language Acquisition: Languages, Contexts, and Learners, Trends in Language Acquisition Research, pages 339–364. John Benjamins Publishing Company.

Colleen E. Morisset, Kathryn E. Barnard, and Cathryn L. Booth. 1995. Toddlers' language development: Sex differences within social risk. Developmental Psychology, 31(5):851–865.

Katherine Nelson. 2006. Narratives from the Crib. Harvard University Press.

Jens Nevens, Jonas Doumen, Paul Van Eecke, and Katrien Beuls. 2022. Language Acquisition through Intention Reading and Pattern Finding. In Proceedings of the 29th International Conference on Computational Linguistics, pages 15–25, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Rochelle S. Newman, Meredith L. Rowe, and Nan Bernstein Ratner. 2016. Input and Uptake at 7 Months Predicts Toddler Vocabulary: The Role of Child-Directed Speech and Infant Processing Skills in Language Development. Journal of Child Language, 43(5):1158–1173.

J. G. Nicholas and A. E. Geers. 1997. Communication of oral deaf and normally hearing children at 36 months of age. Journal of speech, language, and hearing research: JSLHR, 40(6):1314–1327.

Ageliki Nicolopoulou, Hande Ilgaz, Marta Shiro, and Lisa B. Hsin. 2022. "And they had a big, big, very long fight:" The development of evaluative language in preschoolers' oral fictional stories told in a peer-group context. Journal of Child Language, 49(3):522–551.

A. Ninio, C. E. Snow, B. A. Pan, and P. R. Rollins. 1994. Classifying communicative acts in children's interactions. Journal of Communication Disorders, 27(2):157–187.

Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajic, Daniel Hershcovich, Bin Li, Tim O'Gorman, Nianwen Xue, and Daniel Zeman. 2020. MRP 2020: The Second Shared Task on Cross-Framework and Cross-Lingual Meaning Representation Parsing. In Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing, pages 1–22, Online. Association for Computational Linguistics.

Stephan Oepen and Dan Flickinger. 2019. The ERG at MRP 2019: Radically Compositional Semantic Dependencies. In Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning, pages 40–44, Hong Kong. Association for Computational Linguistics.

Stephan Oepen, Kristina Toutanova, Stuart Shieber, Christopher Manning, Dan Flickinger, and Thorsten Brants. 2002. The LinGO Redwoods Treebank: Motivation and Preliminary Applications. In COLING 2002: The 17th International Conference on Computational Linguistics: Project Notes.

Mitsuhiko Ota, Nicola Davies-Jenkins, and Barbora Skarabela. 2018. Why Choo-Choo Is Better Than Train: The Role of Register-Specific Words in Early Vocabulary Growth. Cognitive Science, 42(6):1974–1999.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Fernando Pereira and Yves Schabes. 1992. Inside-Outside Reestimation From Partially Bracketed Corpora. In 30th Annual Meeting of the Association for Computational Linguistics, pages 128–135, Newark, Delaware, USA. Association for Computational Linguistics.

Ann M. Peters. 1987. The role of Imitation in the developing syntax of a blind child. Text - Interdisciplinary Journal for the Study of Discourse, 7(3):289–309.

Geoffrey Pullum and Barbara Scholz. 2001. On the distinction between Generative-Enumerative and Model-Theoretic syntactic frameworks. In Logical Aspects of Computational Linguistics: 4th International Conference (Lecture Notes in Computer Science 2099), pages 17–43. Springer Verlag.

N. B. Ratner. 1984. Patterns of vowel modification in mother-child speech. Journal of Child Language, 11(3):557–578.

E Reiter. 2000. Building natural language generation systems.

Pamela Rosenthal Rollins. 2003. Caregivers' contingent comments to 9-month-old infants: Relationships with later language. Applied psycholinguistics, 24(2):221–234.

Jean Adolphe Rondal. 1976. Maternal speech to normal and Down's syndrome children matched for mean length of utterance. University of Minnesota.

Caroline F. Rowland and Sarah L. Fletcher. 2006. The effect of sampling on estimates of lexical specificity and error rates. Journal of Child Language, 33(4):859–877.

Rachel Rudman and Felicity Titjen. 2018. Language Development. Cambridge University Press.

Jacqueline Sachs. 1983. Talking about the There and Then: The Emergence of Displaced Reference in Parent-Child Discourse. In Children's Language. Psychology Press.

J. R. Saffran, R. N. Aslin, and E. L. Newport. 1996. Statistical learning by 8-month-old infants. Science (New York, N.Y.), 274(5294):1926–1928.

E. Sapir. 1929. The Status of Linguistics as a Science. Language, 5(4):207–214.

R. Keith Sawyer. 2013. Pretend Play As Improvisation: Conversation in the Preschool Classroom. Psychology Press, New York.

Dan Isaac Slobin, Julie Gerhardt, Amy Kyratzis, and Jiansheng Guo. 2014. Social Interaction, Social Context, and Language: Essays in Honor of Susan Ervin-tripp. Psychology Press.

Melanie Soderstrom, Megan Blossom, Rina Foygel, and James L. Morgan. 2008. Acoustical cues and grammatical units in speech to two preverbal infants. Journal of Child Language, 35(4):869–902.

M. Steedman. 1996. Surface Structure and Interpretation. Linguistic Inquiry Monographs. Mit Press.

Mark Steedman. 2000. The syntactic process. MIT Press, Cambridge, MA, USA.

Mark Steedman. 2001. The syntactic process. MIT press.

Anastasia Stoops, M Wu, IHT Jung, and JL Montag. 2024. A novel corpus of naturalistic picture book reading with 2-to 3-year-old children. Language Development Research, 4(1):259–297.

Patrick Suppes. 1974. The semantics of children's language. American Psychologist, 29(2):103–114.

A. L. Theakston, E. V. Lieven, J. M. Pine, and C. F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: An alternative account. Journal of Child Language, 28(1):127–152.

Jodi Tommerdahl and Cynthia D Kilpatrick. 2014. The reliability of morphological analyses in language samples. Language Testing, 31(1):3–18.

V. Valian. 1991. Syntactic subjects in the early speech of American and Italian children. Cognition, 40(1-2):21–81.

Virginia Valian. 2024. Variability in Production is the Rule, Not the Exception, at the Onset of Multi-Word Speech. Language Learning and Development, 20(1):76–78.

Laura Wagner, Sharifa Alghowinhem, Abeer Alwan, Kristina Bowdrie, Cynthia Breazeal, Cynthia G Clopper, Eric Fosler-Lussier, Izabela A Jamsek, Devan Lander, Rajiv Ramnath, et al. 2025. The ohio child speech corpus. Speech Communication, 170:103206.

Amye Warren-Leubecker and John N. Bohannon. 1984. Intonation patterns in child-directed speech: Mother–father differences. Child Development, 55(4):1379–1385.

Sandra R. Waxman and Dana B. Markow. 1995. Words as invitations to form categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology, 29(3):257–302.

Susan Ellis Weismer, Courtney E. Venker, Julia L. Evans, and Maura Jones Moyle. 2013. Fast mapping in late-talking toddlers. Applied Psycholinguistics, 34(1):69–89.

Richard M. Weist, Aleksandra Pawlak, and Karen Hoffman. 2009. Finiteness systems and lexical aspect in child Polish and English. 47(6):1321–1350.

Gordon Wells and C. Gordon Wells. 1981. Learning Through Interaction: Volume 1: The Study of Language Development. Cambridge University Press.

Danielle Wetherell, Nicola Botting, and Gina Conti-Ramsden. 2007. Narrative in adolescent specific language impairment (SLI): A comparison with peers across two different narrative genres. International Journal of Language & Communication Disorders, 42(5):583–605.

Benjamin Lee Whorf. 1956. Language, Thought, and Reality: Selected Writings Of. . . . (Edited by John B. Carroll.). Language, Thought, and Reality: Selected Writings Of. . . . (Edited by John B. Carroll.). Technology Press of MIT, Oxford, England.

Songlin Yang, Yanpeng Zhao, and Kewei Tu. 2021. PCFGs Can Do Better: Inducing Probabilistic Context-Free Grammars with Many Symbols. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1487–1498, Online. Association for Computational Linguistics.

Yajie Ye and Weiwei Sun. 2020. Exact yet Efficient Graph Parsing, Bi-directional Locality and the Constructivist Hypothesis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4100–4110, Online. Association for Computational Linguistics.

Luke S. Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence, UAI'05, pages 658–666, Arlington, Virginia, USA. AUAI Press.

Yuanyuan Zhao, Weiwei Sun, Junjie Cao, and Xiaojun Wan. 2020. Semantic parsing for English as a second language. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6783–6794, Online. Association for Computational Linguistics.

A CHILDES Corpora Used

18 Corpora from the Eng-UK collection from the CHILDES were included. They include Belfast (Henry, 1995), Conti-Ramsden 1 and 4 (Conti-Ramsden and Dykins, 1991; Wetherell et al., 2007), Cuttenden (Cruttenden, 1978), Edinburgh (Ota et al., 2018), Fletcher (Fletcher and Garman, 1988), Forrester (Forrester, 2002), Gathercole/Burns (Gathercole, 1986), Howe (Howe, 1981), KellyQuigley (Kelly et al., 2020), Lara (Rowland and Fletcher, 2006), Manchester (Theakston et al., 2001), MPI-EVA-Manchester (Lieven et al., 2009a), Nuffield (McGillion et al., 2017), Sekali (Morgenstern et al., 2018), Thomas (Lieven et al., 2009b), Tommerdahl (Tommerdahl and Kilpatrick, 2014), and Wells (Wells and Wells, 1981).

61 corpora from the Eng-NA collections were included. They are Bates (Bates et al., 1991), Bernstein Ratner (Ratner, 1984), Bliss (Bliss, 1988), Bloom (Bloom, 1970), Bohannon (Bohannon and Marquis, 1977), Braunwald (Braunwald, 1971), Brent/Siskind (Brent and Siskind, 2001), Brown (Brown, 2013), Champaign (Hadley et al., 2014), Clark (Clark, 1978), Demetras: Trevor and Working (Demetras, 1989), Ellis Weismer (Weismer et al., 2013), Evans, Feldman (Feldman and Menn, 2003), Garvey (Garvey and Hogan, 1973), Gathercole (Gathercole, 1986), Gelman (Gelman et al., 1998), Gillam (Gillam and Pearson, 2004), Gleason (Bellinger and Gleason, 1982), Haggerty (Haggerty, 1930), Hall (Hall et al., 1984), Higginson (Higginson, 1985), HSLLD (Dickinson and Tabors, 2001), Kuczaj (Kuczaj, 1977), MacWhinney (Macwhinney, 2000), McCune (McCune, 1995), McMillan, Morisset (Morisset et al., 1995), Nelson (Nelson, 2006), New England (Ninio et al., 1994), Newman/Ratner (Newman et al., 2016), Nicholas (Nicholas and Geers, 1997), Nicolopoulou (Nicolopoulou et al., 2022), Nippold, OCSC (Wagner et al., 2025), Peters/Wilson (Peters, 1987), POLER (Berl et al., 2005), Post (Demetras et al., 1986), Rollins (Rollins, 2003), Rondal (Rondal, 1976), Sachs (Sachs, 1983), Sawyer (Sawyer, 2013), Snow (MacWhinney and Snow, 1990), Soderstrom (Soderstrom et al., 2008), Sprott (Slobin et al., 2014), StoopsMontag (Stoops et al., 2024), Suppes (Suppes, 1974), Tardif, Valian (Valian, 1991), Van Houten (Houten and J, 1986), Kleeck, Warren (Warren-Leubecker and Bohannon, 1984), and Weist (Weist et al., 2009).