Verb Classification in Child Input: Can Models Learn from Child Input Like Children Do?

Jingying Xu
Michigan State University
xujing21@msu.edu

Molly Thornber
Michigan State University
thornbe5@msu.edu

Abstract

Recent work in NLP has highlighted the question of what models can learn from small, child-like input, in contrast to the scaling-up paradigm that dominates modern language modeling. This study asks whether child-directed speech (CDS) contain enough distributional information to support the learning of fine-grained verb meanings? Using the Brown corpus from CHILDES, we manually annotated 3,000+ utterances containing 19 target verbs for syntactic and semantic cues relevant to telicity (object structure, theme quantization, preposition phrase (PP) type, duration adverbials, aspect marking, etc.). We trained statistical classifiers on (i) linguistic features only, (ii) embedding-only representations, and (iii) hybrid combinations of both. Across both linear (SVM) and non-linear (FFNN) architectures, only Incremental Theme verbs (e.g., *eat*, *built*) were reliably recoverable from CDS; other verb types showed little to no learnable signal. Embeddings increased overall accuracy but did not improve aspectual discrimination, and hybrid models behaved similarly. These results suggest that CDS provides limited distributional evidence for learning most of the verb type distinctions, and that children must rely on additional cues, especially event perception, to acquire the full system.

1 Introduction

Modern language models have been increasingly driven by scaling-up: larger models trained on massive corpora continue to improve across tasks. In contrast, recent initiatives such as BabyLM (Choshen et al., 2024) ask whether a model can learn effectively from small, child-like input. Although BabyLM succeeds in data efficiency rather than child-like cognitive mechanisms, its core goal remains valuable.

A large body of work in language acquisition shows that children *do* rely on statistical distribution in the input. For example, children exploit the distribution of syntactic frames to narrow a verb's interpretation (Gleitman et al., 2005; Gleitman, 1990; Naigles, 1990). Classic findings include: (i) transitives bias causative readings, (ii) intransitives bias activity readings, and (iii) sentential-complement frames bias psych/stative meanings. For instance, infants interpret *he glorped the toy* as "he caused the toy to change," *he glorped* as "he did something himself," and *she wugged that the story was true* as "she thought or knew that the story was true" (Naigles, 1990; Yuan and Fisher, 2009; Hacquard and Lidz, 2018; Perkins et al., 2024). Children do extract statistical patterns from small, naturalistic datasets.

In this sense, computational models trained on child-sized corpora are not meant to mimic children's internal algorithms, but to probe the information content available in the input. Thus, BabyLM's original motivation *should not* be dismissed: if children can learn from sparse, naturalistic speech, then understanding what is learnable from such input using simple statistical learners remains a crucial part of understanding both human acquisition and the limits of small-data NLP.

In the current study, we ask: does the naturalistic child-directed speech (CDS) contain enough distributional information for the model do to recover the finer aspectual meaning of eventive verbs that children eventually master?

We focuses on four classes of verbs:

- **Incremental Theme Verbs.** Verbs such as *eat, drink, build, write* encode a homomorphism (piece-by-piece mapping) between the referent of the direct object and event progress. With *quantized* direct objects (i.e., direct objects that encode a fixed amount of quantity such as *the*

cookie), they pattern as *telic* (i.e., the event entails an inherent endpoint); with *mass/bare plural* objects, they pattern as *atelic* (i.e., the event does not entail an inherent endpoint):

(1) I ate *the cookie in* ten minutes. (telic)

(2) I ate *cookies for* ten minutes. (atelic)

- **Degree-Achievement Verbs.** Verbs such as *dry, clean, widen, cool, change* map events onto gradable scales and freely combine with both *in-/for-X* adverbials even with quantized direct objects:

(3) I cleaned the room *in* ten minutes */for* ten minutes. (telic/atelic)

- **Activity Verbs.** Verbs such as *push, carry, pull* are typically atelic unless *path/result* structure is supplied:

(4) I carried the box *for* ten minutes. (atelic)

(5) I carried the box *to the table in* ten minutes. (telic)

- **Change-of-State Verbs.** Verbs such as *open, break, spill, wake* typically encode a punctual culmination, and are always telic:

(6) I opened the door *in* a second. (telic)

Prior experimental work shows that by age four children already distinguish incremental-theme verbs and change-of-state verbs from activity verbs, and that object properties modulate telicity for incremental-theme verbs in comprehension tasks (Xu and Schmitt, in prep; Ogiela, 2007; Martin et al., 2020). But how are these fine-grained aspectual verb classes learned at the first place? Can they be inferred from distributional cues in child-directed speech (CDS), beyond the coarse syntactic diagnostics of "transitive vs. intransitive vs. sentential complement"?

We hypothesize that verbs cluster by how often they occur with *goal/path PPs*, *particles*, *tense/aspect morphemes*, and *DP object structures* etc. From these distributions, a learner could, in principle, infer each lemma's scale architecture (incremental-theme, degree-scale, path/result, punctual) and thus recover the aspectual categories of verbs that transitivity alone cannot provide.

Research questions In summary, our research questions are:

1. **Recoverability.** To what extent are aspectual verb classes recoverable from the surface distribution of syntactic and semantic cues in child-directed speech (CDS), as probed by simple statistical learners?
2. **Type of cues.** Which types of cues (interpretable syntactic/semantic features versus distributional embeddings) support the recoverability of these aspectual distinctions?
3. **Robustness.** Is the learnability profile robust across different model architectures (e.g., linear vs. non-linear learners), or does it depend on specific modeling assumptions?
4. **Implications for child acquisition**. What do the recoverability patterns imply about the cues that children must rely on?

2 Data and Annotation

2.1 Corpus and Target Verbs

We analyzed the Brown corpus from CHILDES (MacWhinney, 2000), a publicly available repository of child language transcripts widely used in language acquisition research. The Brown corpus contains naturalistic data from three children aged 1;6–5;1 (years;months) (Adam: 2;3–5;2, Eve: 1;6–2;3, Sarah: 2;3–5;1), totaling approximately 136,000 utterances. We focused on the adult-to-child portion of the corpus (about 70,000 utterances, 200,000–250,000 words), from which we extracted verbs with at least 30 attestations across four aspectual categories:

- **Incremental Theme** (INC): *build, draw, drink, eat, write*
- **Degree Achievement** (DGA): *clean, dry, wash*
- **Change-of-State** (COS): *open, close, break, cut, catch*
- **Activity** (ACT): *pull, ride, drive, roll, carry, wipe*

Ungrammatical, idiomatic, and ambiguous uses were excluded. The final dataset contains 3,196 adult-to-child utterances, tiny compared to modern NLP datasets.

2.2 Feature Annotation

Each utterance was manually coded for both syntactic and semantic features relevant to aspectual composition.

Syntactic features

- Theme movement (unaccusatives, passives, wh-questions, relative clauses, tough constructions, etc.)
- Presence of a surface direct object
- Object type (determiner-marked or pronoun vs. bare)
- VP modification by PP adjunct
- Presence of *in-* duration adverbial
- Presence of *for-* duration adverbial
- Aspectual morphology (bare, perfective, progressive)

Semantic features

- Presence of an overt theme (including moved themes)
- Quantization of the theme
- PP semantic role (source, goal, benefactive, path, locative, tool, purpose)

Interestingly, we found **no** instances of overt *for-* or *in-* duration adverbials in CDS. This absence is itself informative: these temporal modifiers, which serve as diagnostics of telicity in adult speech, appear to be rare or entirely absent in child-directed input. Consequently, children's input provides few explicit surface cues that distinguish telic from atelic events. Learners must therefore infer telicity, and, more generally, the aspectual properties of verbs, primarily from the distributional relationships among verbs, objects, prepositions, tense/aspect morphemes and other distributional cues rather than from overt temporal modification.

3 Modeling Framework

3.1 Modeling Design and Feature Representation

Our modeling approach tests whether aspectual verb classes can be inferred from distributional cues in CDS without direct access to the verb lemma. To approximate a learner who has not yet mapped verb forms to stable lexical semantic categories, the target verb lemma is masked in all models. The model must therefore rely on the syntactic and semantic structure of the utterance, rather than memorizing specific verb identities.

Each utterance is represented as a combination of categorical and continuous features:

- Categorical syntactic and semantic features are one-hot encoded.
- Continuous features consist of embedding vectors.

We consider three feature configurations:

1. **Syntactic + semantic features.**

   We extracted 14 manually annotated features for each verb token:

   Syntactic:

   - presence of an overt direct object (`has_DO`),
   - whether the direct object was a pronoun and/or had a determiner (`DO_pron_has_det`),
   - presence of a verb particle (`has_particle`),
   - presence of a prepositional phrase (`has_pp`),
   - progressive marking (`is_progressive`),
   - perfective marking (`is_perfective`).

   Semantic:

   - presence of an overt theme (`theme_present`),
   - whether the theme encodes a fixed quantity (`theme_quantization`),
   - whether prepositional phrases encode source, goal, path, location, tool, or purpose (`source_pp`, `goal_pp`, `path_pp`, `loc_pp`, `tool_pp`, `purpose_pp`).

2. **Embedding features.**

   For the embedding vectors, we use a pretrained DistilBERT tokenizer and model (Sanh et al., 2019) from `HuggingFace`'s `transformers` library to extract embeddings from each utterance. Then, we take only the embeddings of the masked position (i.e. the masked target verb) to be used in the models. The embeddings for each masked token have dimensionality $d_{\text{emb}} = 768$.

3. **Combined features.**

   The hybrid setting concatenates the syntactic, semantic, and embedding features into a single feature vector.

This design tests whether the distributional contexts in CDS alone provide sufficient information for a learner to infer aspectual class, and how much these inferences depend on structured linguistic cues vs. general distributional similarity.

3.2 Model Architectures

We evaluate two types of classification models: SVM and FFNN, which differ in representational capacity but are both standard, well-understood statistical learners. This allows us to test whether the recoverability of aspectual classes from CDS is robust across linear and non-linear architectures. [1]

- **Support vector classifier** (from `scikit-learn`'s `SVC` model) with the RBF kernel and a scaled kernel coefficient γ. The SVM model was chosen because of its ability to learn from both low- and high-dimension feature vectors, meaning it should be able to perform well on the semantic and semantic features in addition to the embedding features.
- **Feed-forward neural network** (from `scikit-learn`'s `MLPClassifier` model) with 5 hidden layers of sizes 128, 128, 64, 32, and 16, the L-BFGS solver, L2 regularization term of $\alpha = 0.0001$, and a maximum of 200 iterations. The L-BFGS solver was chosen since the dataset was relatively small, and the rest of the hyperparameters were selected after testing multiple combinations for optimal performance. The FFNN model was chosen because it is able to handle the integration of syntactic, semantic, and embedding features, as well as non-linear patterns.

4 Results

Each model is evaluated using 5-fold cross-validation, stratified by aspectual class (label) and grouping by the lemma of the masked verb. The true labels y_i and predicted labels $\hat{y}_i$ are collected across the folds, and metrics are calculated on the compiled label data.

Models were trained on the gated features, including sentences with movement or passives. For each model, we additionally trained a one-vs-rest (OVR) classifier, using `scikit-learn`'s `OneVsRestClassifier` model wrapper, but since the OVR performance was nearly identical to the multiclass setting and did not change the learnability pattern, we report only the multiclass results.

4.1 Linguistic Features Only

Tables 1-2 show results from SVM and FFNN classifiers trained on manually annotated syntactic and semantic features.

Both models exhibit the same qualitative pattern: only Incremental Theme (INC) verbs are reliably recoverable, with F1 $\approx$ 0.59–0.60. The remaining classes, Activity (ACT), Change-of-State (COS), and Degree Achievement (DGA) verbs, yield very low or near-zero F1 scores. This indicates that CDS contains robust surface cues for only Incremental Theme verbs while providing insufficient distributional evidence for other aspectual classes.

SVM	Prec.	Recall	F1	Support
ACT	0.1559	0.0605	0.0872	479
COS	0.2024	0.2388	0.2191	779
DGA	0.0000	0.0000	0.0000	148
INC	**0.5534**	**0.6555**	**0.6001**	1605
Accuracy			0.4208	3011
Macro	0.2279	0.2387	0.2266	3011
Weighted	0.3721	0.4208	0.3904	3011

Table 1: SVM multiclass classification performance (syntactic + semantic features).

4.2 Embedding Features Only

Next, we evaluate models trained only on embeddings. As shown in Tables 3-4, both SVM and FFNN achieve higher overall accuracy than in the linguistically constrained condition (SVM: 0.49; FFNN: 0.48), but the class-level pattern reveals that this improvement does not reflect better aspectual discrimination. The same asymmetry persists:

[1] We also implemented logistic regression, decision trees, and an RNN as additional baselines. Across all models, the qualitative pattern was identical: only incremental-theme verbs were reliably recoverable from CDS. We report SVM as the best-performing linear model and FFNN as the best-performing non-linear model here.

FFNN	Prec.	Recall	F1	Support
ACT	0.1659	0.0710	0.0994	479
COS	0.2004	0.2401	0.2185	779
DGA	0.0000	0.0000	0.0000	148
INC	**0.5528**	**0.6424**	**0.5942**	1605
Accuracy			0.4158	3011
Macro	0.2298	0.2384	0.2280	3011
Weighted	0.3729	0.4158	0.3891	3011

Table 2: FFNN multiclass classification performance (syntactic + semantic features).

INC verbs achieve the highest F1 ($\approx$ 0.66-0.69). The rest all shows weak, inconsistent recoverability. This suggests that distributional meaning alone is insufficient to recover most aspectual verb classes beyond incremental-theme verbs.

SVM	Prec.	Recall	F1	Support
ACT	0.1037	0.0710	0.0843	479
COS	0.3020	0.2683	0.2842	779
DGA	0.3636	0.0541	0.0941	148
INC	**0.6227**	**0.7639**	**0.6861**	1605
Accuracy			0.4905	3011
Macro	0.3480	0.2893	0.2872	3011
Weighted	0.4444	0.4905	0.4573	3011

Table 3: SVM multiclass classification performance (embeddings only).

FFNN	Prec.	Recall	F1	Support
ACT	0.1498	0.1608	0.1551	479
COS	0.3636	0.3646	0.3641	779
DGA	0.1987	0.2027	0.2007	148
INC	**0.6722**	**0.6555**	**0.6637**	1605
Accuracy			0.4792	3011
Macro	0.3461	0.3459	0.3459	3011
Weighted	0.4860	0.4792	0.4825	3011

Table 4: FFNN multiclass classification performance (embeddings only).

4.3 Linguistic and Embedding features

Finally, we combine both feature types: annotated linguistic cues and distributional embeddings. If the two sources of information encoded complementary aspects of aspectual structure, we would expect improved performance in this hybrid setting. However, the results (Tables 5-6) show that the hybrid models behave almost identically to the embeddings-only models. Weighted accuracy increases slightly, but aspectual distinctions do not become more learnable when structured linguistic cues are added. Incremental Theme verbs remain the only class with consistently high F1 scores (SVM: 0.69, FFNN: 0.65). The remaining classes show low-to-moderate performance.

This indicates that structured linguistic information does not interact with distributional cues to enhance aspectual discrimination. The learnability asymmetry might thus a property of the CDS input, not of the modeling architecture.

SVM	Prec.	Recall	F1	Support
ACT	0.1051	0.0731	0.0862	479
COS	0.3004	0.2657	0.2820	779
DGA	0.3810	0.0541	0.0947	148
INC	**0.6230**	**0.7639**	**0.6863**	1605
Accuracy			0.4902	3011
Macro	0.3524	0.2892	0.2873	3011
Weighted	0.4452	0.4902	0.4571	3011

Table 5: SVM multiclass classification performance (syntactic + semantic + embedding features).

FFNN	Prec.	Recall	F1	Support
ACT	0.1434	0.1503	0.1468	479
COS	0.3178	0.3338	0.3256	779
DGA	0.1871	0.2162	0.2006	148
INC	**0.6664**	**0.6312**	**0.6483**	1605
Accuracy			0.4573	3011
Macro	0.3287	0.3329	0.3303	3011
Weighted	0.4695	0.4573	0.4630	3011

Table 6: FFNN multiclass classification performance (syntactic + semantic + embedding features).

4.4 Identifying Crucial Cues for Learning Incremental Theme Verbs

To determine which features drive the models' success on Incremental Theme (INC) verbs, we conducted an additional one-vs-rest (OVR) analysis using only two gated features:

1. `has_DO`: whether the verb appears with an overt direct object

2. DO_det: when an object appears, whether it bears a determiner

Both SVM and FFNN models continued to classify INC verbs with high reliability (Tables 7-8), achieving recall of 0.85 and F1 of 0.66 for incremental verbs, with F1 = 0.00 for all other classes.

SVM	Prec.	Recall	F1	Support
ACT	—	0.0000	0.0000	479
COS	—	0.0000	0.0000	779
DGA	—	0.0000	0.0000	148
INC	**0.5330**	**1.0000**	**0.6954**	1605
Accuracy			0.5330	3011
Macro	0.5330	0.2500	0.1654	3011
Weighted	0.5330	0.5330	0.3707	3011

Table 7: SVM one-vs-rest classification using only has_DO and DO_det.

FFNN	Prec.	Recall	F1	Support
ACT	—	0.0000	0.0000	479
COS	0.0000	0.0000	0.0000	779
DGA	—	0.0000	0.0000	148
INC	**0.5422**	**0.8486**	**0.6616**	1605
Accuracy			0.4523	3011
Macro	0.2711	0.2121	0.1654	3011
Weighted	0.3650	0.4523	0.3527	3011

Table 8: FFNN one-vs-rest classification using only has_DO and DO_det.

The raw corpus distributions reveal the source of this separability:

1. INC verbs omit objects much more often than the rest verb classes (55.8% vs. 76–84%).

2. When objects appear, determiners occur less frequently with INC verbs than the rest verb classes (75% vs. 88–93%).

These findings indicate that has_DO and DO_det alone provide nearly all of the discriminative information that the larger models used to recover incremental-theme verbs. In other words, the robust learnability of this class arises from very specific distributional cues related to object realization.

5 Discussion

5.1 Learnability Asymmetry

The present study asked whether aspectual verb classes can be recovered from the distribution of syntactic, semantic, and lexical cues in child-directed speech (CDS). Across all feature sets and architectures, the results reveal a striking and highly consistent asymmetry: Incremental Theme (INC) verbs are reliably learnable from CDS, whereas Activity (ACT), Change-of-State (COS), and Degree Achievement (DGA) verbs are not.

This pattern holds for linguistic-only models, embedding-only models, and hybrid models that combine both feature types, and for both linear (SVM) and nonlinear (FFNN) learners. This indicates that the learnability pattern is not model-specific, but rather reflects how aspectual information is (or is not) encoded in CDS. In other words, the distributional footprint of aspectual classes is uneven: only incremental-theme verbs leave statistically robust surface cues.

Our gated-feature analysis clarifies why INC verbs are learnable. When models had access only to has_DO and DO_det, they still achieved high recall and F1 for incremental verbs and failed completely on all other classes. This behavior mirrors the class-level asymmetry in the full models. The raw distributions show that incremental-theme verbs are unique in two ways: they take expressed objects less often overall, and when they do, those objects less often bear determiners. These two object-realization cues capture nearly all of the signal that separates incremental-theme verbs from the rest.

5.2 Model Success Does Not Imply Child Mechanisms

Although the distributional cues for INC verbs are strong, it does not follow that children track them. Children are unlikely to compute corpus-level distinctions such as "Verb X occurs with an overt object 20% less often than Verb Y," or "this class shows determiners 15% less frequently." Children do not monitor corpus-level frequency differences at this level of granularity.

Thus, model success for INC verbs should be di-

rectly interpreted as evidence that CDS contains information sufficient in principle, not that children rely on the same cues. The failure of models to distinguish most aspectual classes suggests that children must draw on non-distributional sources of evidence, such as event perception in the real world and constructions that encode clear event boundaries such as *finish X* and *stop X*.

5.3 Limitation and Future Direction

A central limitation of this study is the small size of the annotated dataset. Although the Brown corpus offers rich, high-quality transcripts, the number of tokens for each verb class remains limited, especially for Degree Achievement verbs. Larger corpora will be necessary to determine whether the learnability asymmetry observed here generalizes, especially across more balanced datasets.

A second limitation is that our models access only textual CDS, with no visual or situational context. In real acquisition, children learn verb meanings from multimodal input: they hear language while simultaneously perceiving events. Thus, models trained on text alone represent only a fraction of the information available to human learners.

Multimodal learning therefore represents a promising direction for future work. A key question is whether visually grounded models can recover the aspectual distinctions that are not learnable from CDS alone. Future studies should test whether such models (i) predict event completion differently across verb classes, such as expecting a completed endpoint for *eat the apple* but not for *push the cart*; (ii) distinguish telic–atelic minimal pairs such as *eat the cookie* vs. *eat cookies* or *push the cart to the corner* vs. *push the cart*; (iii) show context-dependent flexibility in judging telic descriptions of incomplete events, such as interpreting *He ate three cookies* as true in situations where a character takes a bite of each of the three cookies, but false when the character eats two and a half cookies (patterns that humans show but that purely distributional learners cannot capture (Xu and Schmitt, 2025)).

6 Conclusion

Our results show that child-directed speech contains only limited distributional evidence for learning aspectual verb meanings. Across models and feature sets, only incremental-theme verbs were reliably recoverable, largely due to two simple object-realization cues (object omission and determiner patterns). Other aspectual classes left no consistent surface signal. These findings suggest that while CDS supports learning for a narrow subset of verbs, children must rely on additional sources of information.

7 Work Division

All tasks of annotation, modeling, and writing are evenly divided between the authors.

References

Leshem Choshen et al. 2024. The BabyLM Challenge 2024: Sample-efficient pretraining on a developmentally plausible corpus. In *BabyLM Workshop*.

Lila Gleitman. 1990. The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Lila Gleitman, Kimberly Cassidy, Rebecca Nappa, Anna Papafragou, and John C. Trueswell. 2005. Hard words. *Language Learning and Development*, 1(1):23–64.

Valentine Hacquard and Jeffrey Lidz. 2018. Bootstrapping structural and conceptual knowledge in the acquisition of attitude verbs. *Annual Review of Linguistics*.

Brian MacWhinney. 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd edition. Lawrence Erlbaum Associates, Mahwah, NJ.

Fabienne Martin, Hamida Demirdache, Irene Garcia del Real, and Nina Kazanina. 2020. Children's non-adultlike interpretations of telic predicates across languages. *Linguistics*, 58(5):1161–1202.

Letitia R. Naigles. 1990. Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2):357–374.

Diane Ogiela. 2007. *Development of Telicity Interpretation: Sensitivity to Verb-Type and Determiner-Type*. Ph.D. thesis, Michigan State University.

Laurel Perkins, Tyler Knowlton, Alexander Williams, and Jeffrey Lidz. 2024. Thematic content, not number matching, drives syntactic bootstrapping. *Language Learning and Development*.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Jingying Xu and Cristina Schmitt. 2025. Pragmatic accommodation in judging event culmination. In *Proceedings of the 34th Semantics and Linguistic Theory (SALT 34)*.

Jingying Xu and Cristina Schmitt. in prep. Knowing the instructions, struggling to implement: Telicity judgments in Mandarin-speaking children. Manuscript in preparation.

Sylvia Yuan and Cynthia Fisher. 2009. "really? she blicked the baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*.