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

Jingying Xu & Molly Thornber

CSE842

December 1, 2025

Overview

Introduction

Fill the blank with the prepositions: *in*/*for*:

(1) a. She ate the cookie ___ ten minutes.
    b. She cleaned the room ___ ten minutes.
    c. She carried the box ___ ten minutes.
    d. She opened the door ___ a second.

Introduction

These examples show **four different types of verbs**, each describing a different kind of event. Some verbs naturally point to an **endpoint** (something is finished), while others describe **ongoing activities** with no built-in finish.

(2)
a. She ate the cookie **in** ten minutes.
   She ate cookies **for** ten minutes.
b. She cleaned the room **in/for** ten minutes.
c. She carried the box **for** ten minutes.
   She carried the box to the table **in** ten minutes.
d. She opened the door **in** a second.

Introduction

Prior language acquisition literature shows that **children by age 4** already distinguish those verb types in comprehension tasks (though not without debate) (Ogiela, 2007; Xu and Schmitt, prep; Martin et al., 2020).

Question: What in the input makes this learning possible?

From Children to Models

Computational models trained on child corpora serve as:

- **Probes** of the information available in the input, and
- **Tools** for testing what can be learned from that input,

not as models of children's cognitive mechanisms.

Research Questions

Does the child input contain enough distributional information for the model to recover the four classes of verbs that children eventually master?

- **Recoverability:** Can models recover the verb classes from naturalistic adult-to-child speech?
- **Cues:** Which cues matter most: specific hand-annotated linguistic features or distributional embeddings?
- **Robustness:** Do these results hold across different model architectures (linear vs. non-linear)?
- **Acquisition:** What do these learnability patterns reveal about the cues children may rely on?

Dataset

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). We focused on the **adult-to-child** utterances (about 70,000 utterances), from which we extracted verbs with at least 30 attestations across four categories:

- ***eat*-type**: *build, draw, drink, eat, write*
- ***clean*-type**: *clean, dry, wash*
- ***open*-type**: *open, close, break, cut, catch*
- ***carry*-type**: *pull, ride, drive, roll, carry, wipe*

Ungrammatical, idiomatic, and ambiguous uses were excluded. The final dataset contains **3,196 adult-to-child utterances**.

Feature Annotation

Each utterance was manually coded for both syntactic and semantic features:

- Syntactic features
  - Presence of a direct object
  - Presence of a determiner with the direct object (if there is one)
  - Presence of a verb particle
  - Presence of a prepositional phrase
  - Presence of *in*- duration adverbial
  - Presence of *for*- duration adverbial
  - Presence progressive verb marking
  - Presence of past/perfective verb marking
- Semantic features
  - Presence of an overt theme
  - Whether the theme (if there is one) encodes a fixed quantity
  - Whether prepositional phrases (if there is one) encode source, goal, path, location, tool, or purpose

Model Design: Features

1. **Linguistic features**: Include 14 manually annotated syntactic and semantic features for each verb token.
2. **Embedding features only**: 768-dimensional DistilBERT (Sanh et al., 2019) token embeddings for the masked target lemma, learned during training, representing distributional co-occurrence patterns without explicit structural encoding.
3. **Combined features**: Integrates linguistic and embedding features.

Model Design: Classifiers

1. **Support vector classifier** (SVM model)
   - `scikit-learn`'s `SVC` model
   - RBF kernel with scaled γ
   - Can handle low- and high-dimensional feature vectors – should perform well on linguistic features as well as embeddings
2. **Multilayer perceptron classifier** (FFNN model)
   - `scikit-learn`'s `MLPClassifier` model
   - Five hidden layers of sizes 128, 128, 64, 32, and 16
   - L-BFGS solver (good for small datasets)
   - L2 regularization term $\alpha = 0.0001$
   - Maximum of 200 iterations
   - Can handle the integration of multiple feature types and non-linear patterns

Models were run with 5-fold cross-validation on both multi-class classification and one-vs-rest (OVR) classification.

Results: Overall summaries

		Multi-Class			OVR		
Model	**Features**	**Prec.**	**Recall**	**F1**	**Prec.**	**Recall**	**F1**
FFNN	Linguistic	0.2298	0.2384	0.2280	0.2265	0.2366	0.2256
SVM	Linguistic	0.2279	0.2387	0.2266	0.2441	0.2450	0.2398
FFNN	Embedding	0.3461	**0.3459**	**0.3459**	0.3560	**0.3559**	**0.3559**
SVM	Embedding	0.3480	0.2893	0.2872	**0.3578**	0.3054	0.3091
FFNN	Combined	0.3287	0.3329	0.3303	0.3503	0.3523	0.3512
SVM	Combined	**0.3524**	0.2892	0.2873	0.3572	0.3050	0.3092

Table: Masked classification performance on different feature combinations (support = 3011)

Results: F1 scores per-label

Model	Features	carry	open	clean	eat
Support		479	779	148	1605
FFNN	Linguistic	0.0886	0.2187	0.0000	0.5952
SVM	Linguistic	0.0669	0.2755	0.0350	0.5816
FFNN	Embedding	0.1491	**0.3430**	**0.2416**	0.6901
SVM	Embedding	0.0892	0.3055	0.1505	**0.6912**
FFNN	Combined	**0.1645**	0.3378	0.2361	0.6665
SVM	Combined	0.0846	0.3026	0.1596	0.6901

Table: F1 scores for masked token one-vs-rest classification per label (multi-class results were very similar but generally slightly worse)

Results: Crucial cues for *eat*-type verbs

Which features drive success on *eat*-type verbs?

- `has_DO`: whether the verb appears with an overt direct object (i.e. "eat the cookies")
- `DO_det`: when an object appears, whether it bears a determiner (i.e. "eat the cookies")

We ran an additional OVR analysis using only these two features:

Model	**carry**	**open**	**clean**	**eat**
Support	479	779	148	1605
FFNN	0.0000	0.0000	0.0000	0.6616
SVM	0.0000	0.0000	0.0350	0.6954

Table: F1 scores for masked token one-vs-rest classification per label using only two features

Both models still classify *eat*-type verbs with high reliability.

Findings

- Linear (SVM) and non-linear (FFNN) models perform comparably well – pattern is likely inherent to data, not model
- Embeddings boost overall accuracy but don't improve aspectual discrimination
- Combined features perform about the same as embedding-only features
- *eat*-type have by far the best performance, but are also the most represented in the training data
    1. Option 1: Adult-to-child utterances doe not contain enough statistical cues to learn aspectual classes other than *eat*-type
    2. Option 2: The data has too few examples of the other verb types for the models to learn these classes
- Although the distributional cues for *eat*-type verbs are strong, it does not follow that children track them.

Thanks!

Questions?

References I

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

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

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

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Xu, J. and Schmitt, C. (in prep). Knowing the instructions, struggling to implement: Telicity judgments in Mandarin-speaking children. Manuscript in preparation.