

# Modeling Language Learning from the Ground Up

Christos Christodoulopoulos

Amazon Research Cambridge  
chrchrs@amazon.co.uk

Dan Roth

University of Pennsylvania  
danroth@seas.upenn.edu

Cynthia Fisher

University of Illinois Urbana-Champaign  
clfishe@illinois.edu

## Introduction

- Language learners must map lexical units to underlying semantic representations
- **Syntactic bootstrapping:**
  - learners use syntactic structure to acquire the meanings of novel verbs
- **Structure Mapping:**
  - children innately expect a 1-to-1 mapping from nouns to semantic arguments
  - they use this information to identify verbs and assign semantic roles to their arguments
- **This work:**
  - How can syntactic bootstrapping arise?
  - Does identifying verbs help the learner to identify semantic roles?
  - Are 'seed' verbs needed for verb identification?

### Assumptions:

- 1) there exist verbs
- 2) nouns are not verbs
- 3) once a verb, always a verb

### Seed nouns:

daddy, baby, eyes, you, I, ...

The girl **krads** the boy

Features

**Lexical:** the word tokens of the noun and verb  
girl, krads

**Noun Pattern:** the order of the noun relative to other nouns  
1st of 2

**Verb Position:** the position of the noun relative to the verb  
before

syntax

semantics

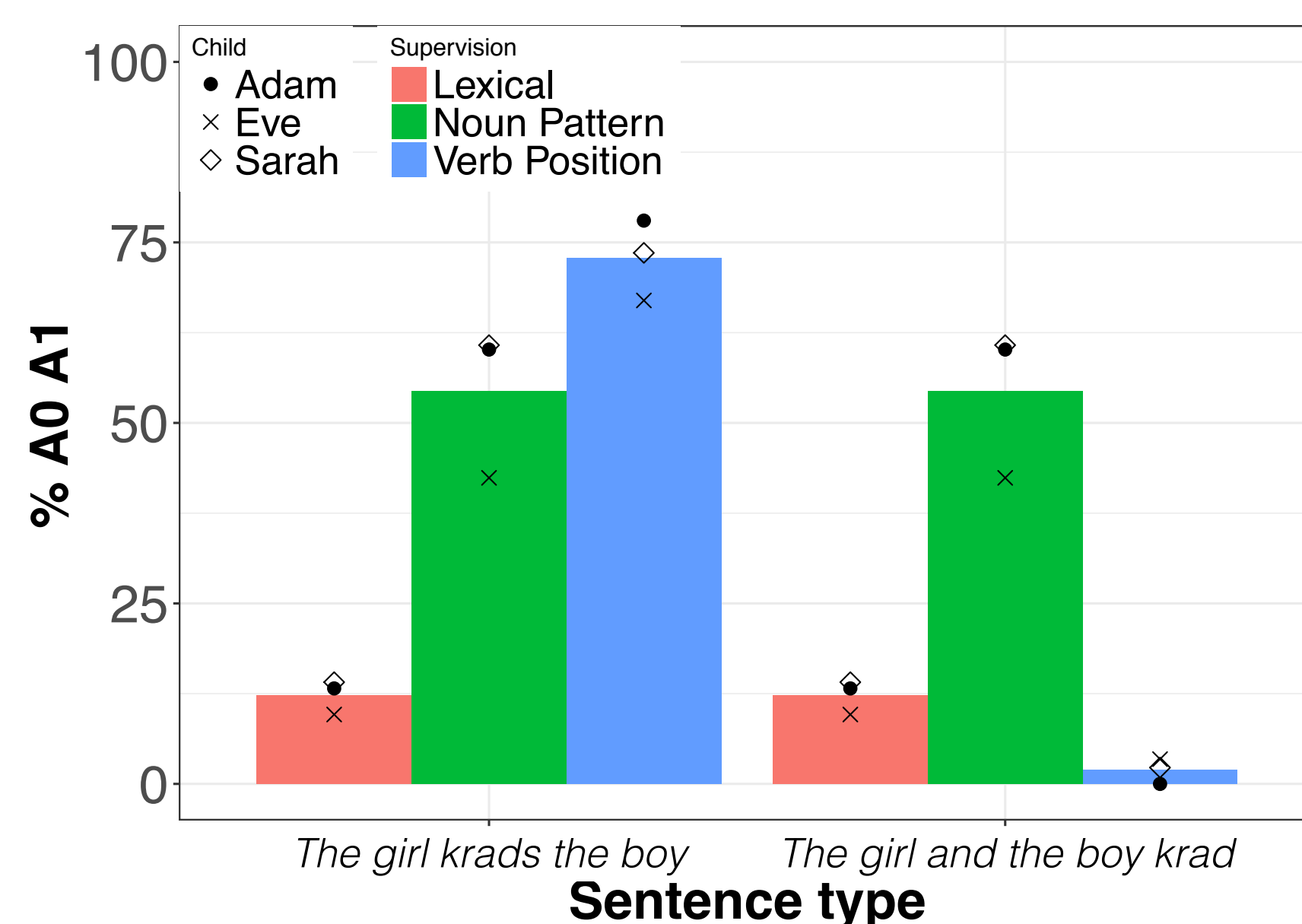
Predictions

girl: agent  
boy: patient

girl: Animate  
boy: Animate

## Experiment 1

Are simple linear syntactic features sufficient to learn to understand simple sentences?



### Setting

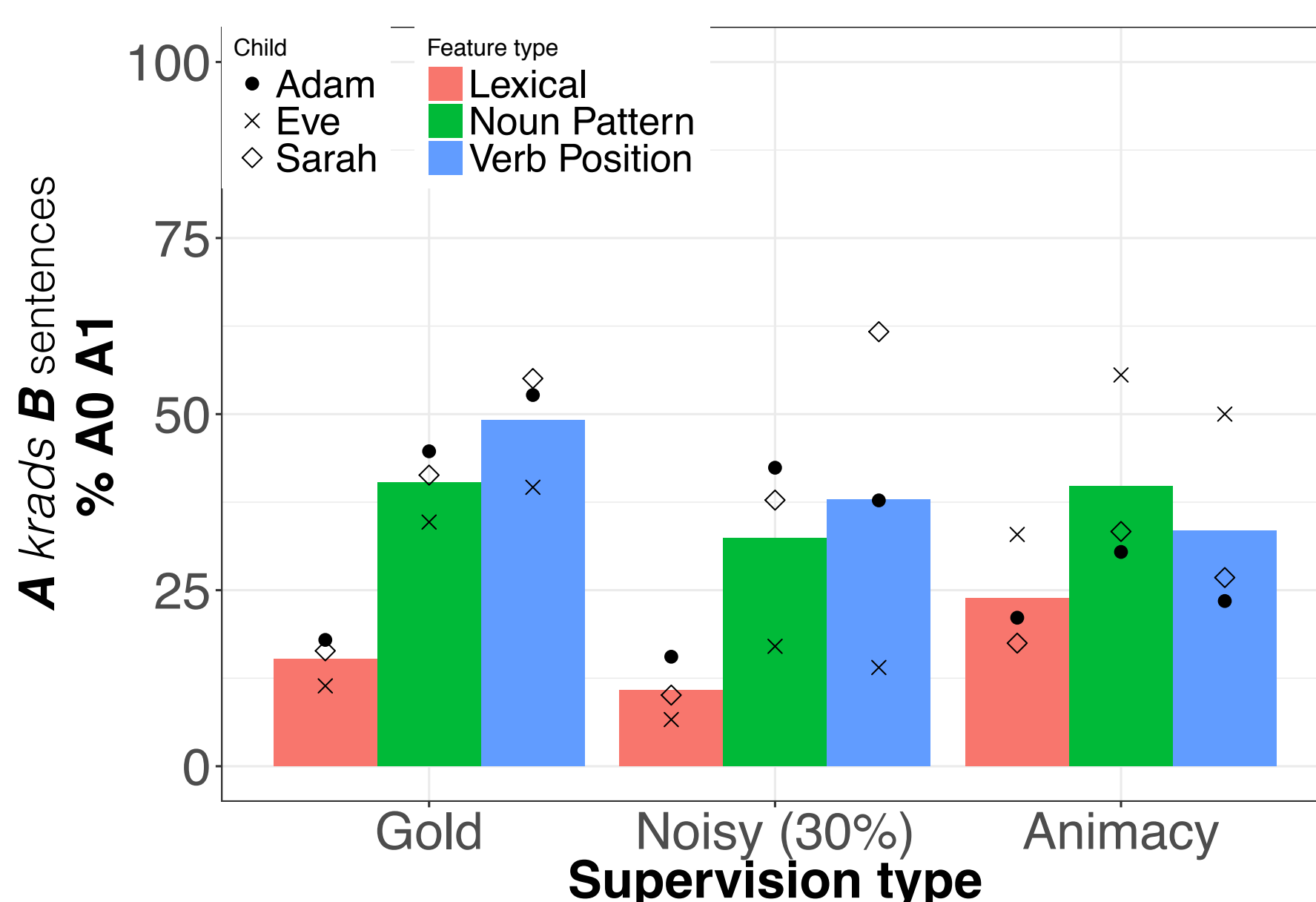
- Veridical semantic feedback

### Findings

- **Noun Pattern** features enough to learn Agent/Patient roles
  - But cause errors with conjoined-subject intransitive sentences
  - Similar to children in Gertner & Fisher (2012)
- Verb position features eliminate errors with conjoined-subject sentences

## Experiment 2

What happens if the semantic feedback is not veridical?



### Setting

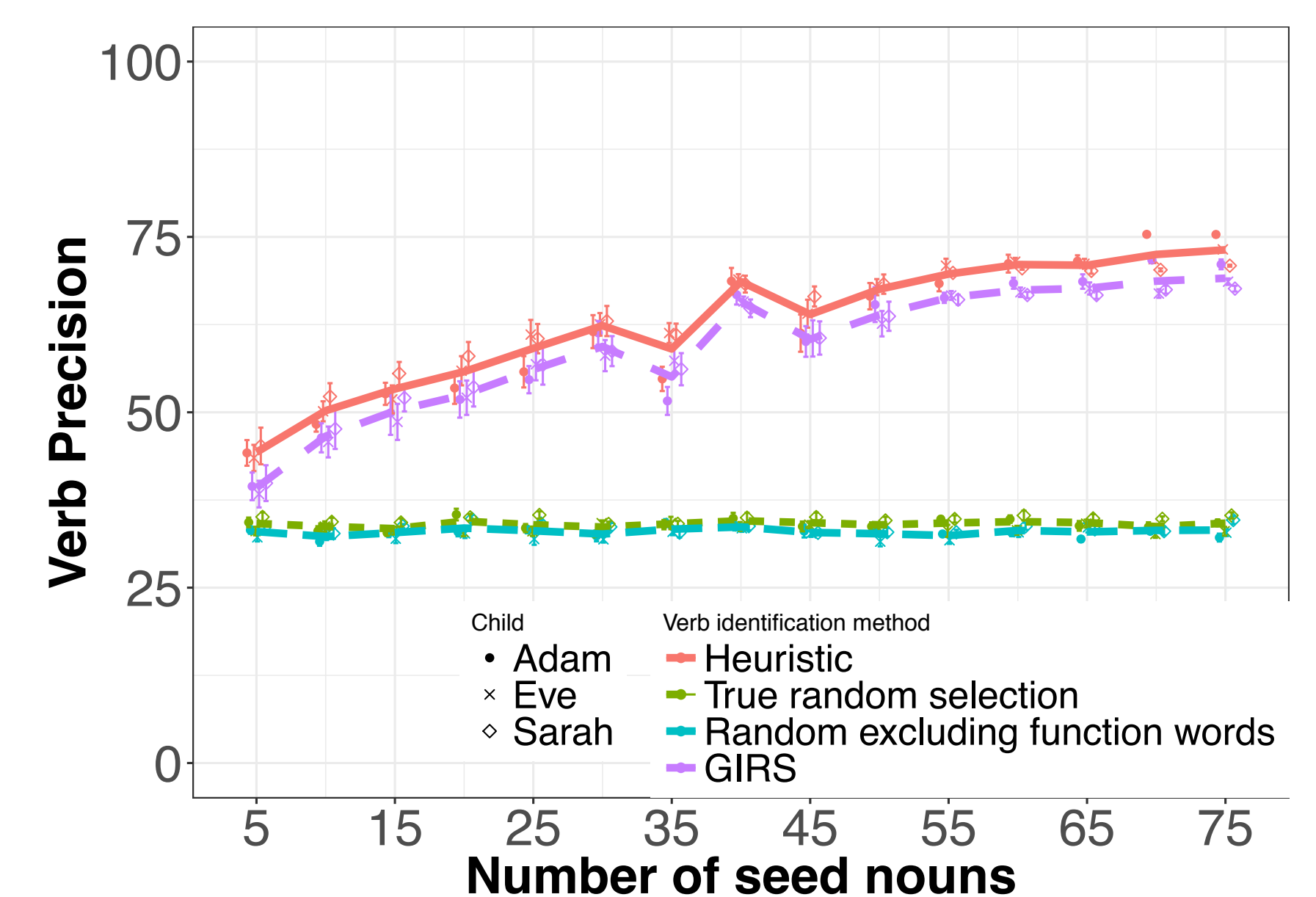
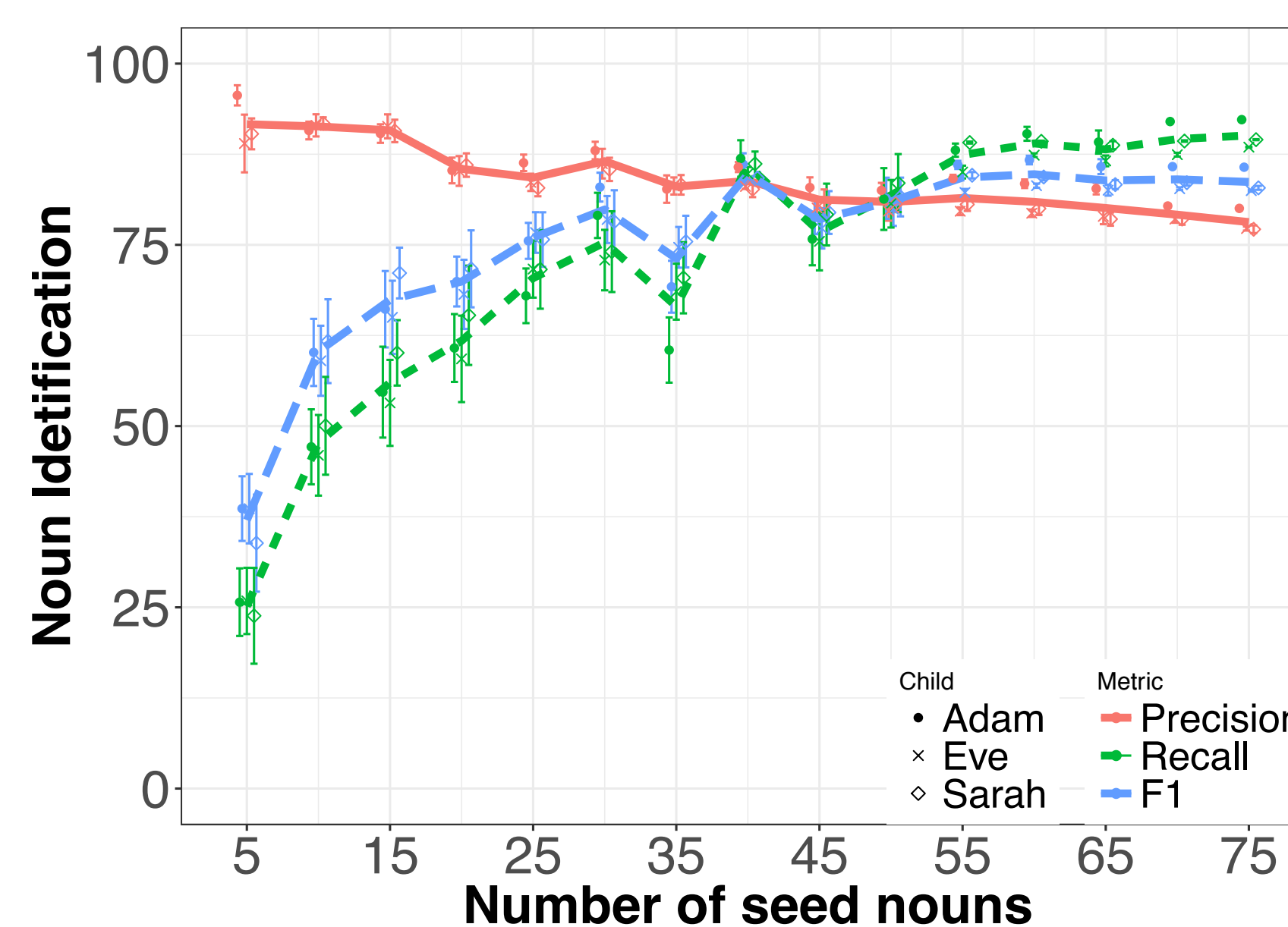
- Identifying nouns and verbs using Experiment 2 process
- Different levels of semantic feedback

### Findings

- Agent/Patient roles learned using noun/verb features even in the presence of significant noise

## Experiment 2

Can nouns nouns and verbs be identified without syntactic feedback?



### Setting

- HMM + 'seed' nouns for noun identification
- Verb identification using:
  - **Heuristic** based on Structure-Mapping
  - **GIRS**: Grammatically-informed random search
- Veridical semantic feedback

### Findings

- Verbs can be identified without any 'seed' verbs
- Random guessing is as good as heuristic iff:
  - a sentence should contain a verb
  - it can't also be a noun or a function word

