UNIVERSITY OF GRONINGEN
InCLoW Reading Group

PRESENTED BY
Francesca Padovani

DATE
12/06/2025

MODELING CHILDREN'S GRAMMAR LEARNING VIA CAREGIVER FEEDBACK IN NATURAL CONVERSATIONS

Mitja Nikolaus & Abdellah Fourtassi, 2025
(Université de Toulouse - Aix Marseille Université)

Glimpse on the work by Stöpler et al. 2025
Towards Developmentally Plausible Rewards: Communicative Success as a Learning Signal for Interactive Language Models

01 THE ROLE OF NEGATIVE EVIDENCE

→ Children are not merely passive recipients of input, **social interaction** shapes their learning experience

→ Children’s actions, vocalizations, and emerging language often **elicit contingent feedback** from caregivers

→ **Syntax learning** (focus of this study) **-** if children produce ungrammatical utterances, caregivers can respond with **distinctive forms of feedback** (e.g., clarification requests), signaling that the child’s utterance was problematic

Well-known case of overgeneralization of regular verbs forms to irregular verbs

👶 “He go-ed to the park.”

👩 “You mean he went to the park?”

Frequent subject and object omission in children’s speech

👶 “Don’t want to go.”

👩 “Who doesn’t want to go?”

01

LANGUAGE ACQUISITION HYPOTESIS

Can caregiver feedback help children learn grammar? (debates on **grammar learnability** and **innateness**)

Empirically addressing this issue REQUIRES analyzing children's natural learning environments and testing 2 key hypotheses:

*1. Parents provide **negative evidence that is contingent on children's grammatical errors** (Brown & Hanlon, 1970, Nikolaus, Prévot, & Fourtassi, 2023)*

*2. Such **negative evidence provides the child with learning gains** in grammar beyond what can be induced from the input.*

→ Measuring the immediate effect of caregiver's feedback, e.g., examining if the child's turn following feedback improves in grammaticality. However, **immediate responses** are not necessarily indicative of a permanent change in children's grammatical knowledge.

(Chouinard & Clark, 2003; Nikolaus et al., 2023, Valian, 1998)

→ Also **longitudinal studies** have been proposed wrt the testing of this hypothesis, and the results are mixed.

(Proctor-Williams, Fey, & Loeb, 2001, Morgan, Bonamo, & Travis, 1995).

01

ADDRESSING THIS GAP

Intervention/experimental designs can be understood as addressing this gap, thanks to randomized controls.
(Kulinich, Royle, & Valois, 2019)

Prior studies often rely on **artificial settings** that lack real-world interaction.

This consideration is of utmost importance here, as the goal is not only to test whether children can learn, but also to demonstrate that learning gains are in principle possible from actual caregiver feedback in natural interaction. This is the gap that the current study aims to address.

CURRENT STUDY

- **Language Modeling** + fine-tuning with **Reinforcement Learning**
- First work to **move beyond the child's input** and **examine the role of natural feedback** contingent on children's own production
- Demonstrating that it allows to address the question of learning gains from the caregiver's contingent feedback in a **controlled** and **ecological way**

RELATED WORKS

- **A key distinction** with previous works is that they all **relied on artificial reward signals** and/or **artificial proxies** for **caregiver feedback**
  (e.g., a virtual agent or a pre-trained Large Language Model (LLM) playing the caregiver's role)
- This study test learnability from the actual social feedback that children **receive from caregivers in everyday interactions**, thereby moving beyond proof-of-concept simulations toward modeling that can support empirical conclusions.

02 FOCUS ON CLARIFICATION REQUESTS

Utterance (Child): *Need some milk* [subject omission error]
Response (Caregiver): *Hm?* [clarification request]
Follow-up (Child): *I need some milk.* [child corrects the error]

— EllisWeismer corpus, LT/42pc/22175.cha (Heilmann, Weismer, Evans, & Hollar, 2005)

CR were found to be a **common feature of human communication** across many cultures (Dingemanse et al., 2015; Dingemanse, Torreira, & Enfield, 2013)

Does not depend on a specific parenting style or on a desire to correct the interlocutor's language; rather, it is a **general mechanism to repair a breakdown in communication** (Nikolaus & Fourtassi, 2023)

02 GOAL TO ADDRESS THE 2 HYPOTHESIS

H1

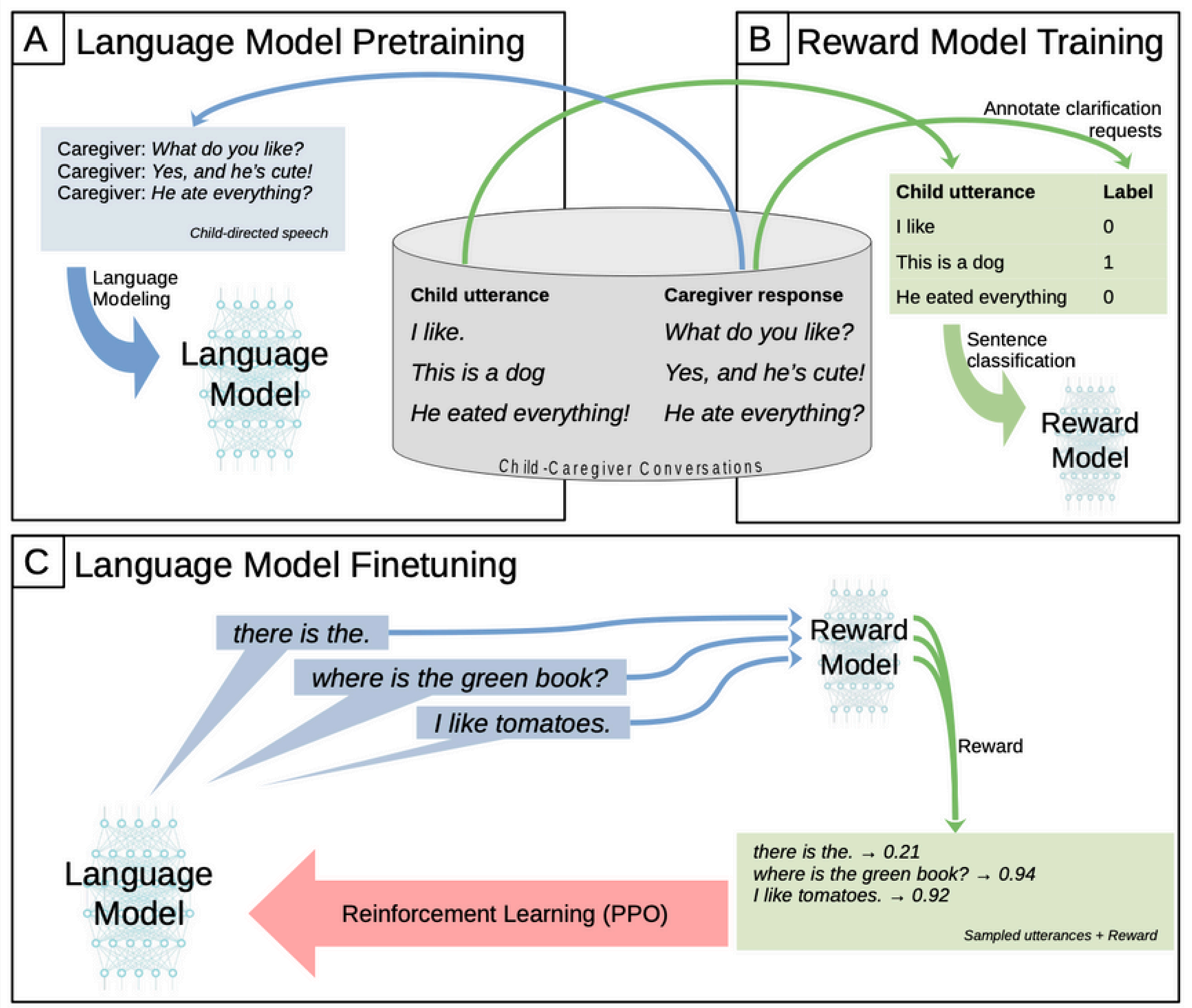
Annotation of **grammaticality** with an **existing tool** for automatic annotation specifically developed for child-caregiver conversations (Nikolaus et al., 2024).
To automatically **annotate clarification requests** in such conversations, a **new tool** is developed.

H2

FIRST: **train a baseline Language Model** (a customized version of GPT2) on child-directed speech, excluding children's utterances (Panel A)

SECOND: **train a reward model to learn** – from pairs of child and parent utterances – **to generalize which utterances tend to trigger CR** from caregivers (Panel B)

THIRD: **use the reward model to fine-tune the input-based language models** using reinforcement learning (Panel C), which operationalizes learning from negative evidence



The reward model learn to assign a reward value of 0 to utterances that elicit a clarification request, and a value of 1 to those that do not

03 METHODS - setup

H1

Examine **relationship between children's ungrammatical utterances and caregiver's CR** *at a large scale*, using English-language CHILDES (MacWhinney, 2014). Data from 1,128 children (10-60 months). Total of 475,237 utterances with their corresponding responses from the caregiver.

- Tool to assign **grammaticality judgments** in child-caregiver *dialog. It* classifies each child utterance as either *ungrammatical*, *grammatical*, or *ambiguous*.
- To **annotate CRs**, they train a new automatic tagger + fine-tune *deberta-v3-xsmall* on *classifying utterances into CRs or otherwise*, using N=1000 manually annotated clarification requests from caregivers' data. This classifier reach an accuracy of 0.92 on a held-out test set (20% of the data).

H2

LANGUAGE MODEL PRE-TRAINING (input-based baseline used to determine whether CR improves learning above and beyond the input)

To test the **effect of data size**, they train three baseline models on increasing amounts of data input: **0.1M**, **1M**, and **10M words**.

3 models with different seeds and the best model for each run is selected based on the loss on a held-out validation set (10% of the data)

hidden layers	2
attention heads	8
hidden layer size	512
simple-word level tokenizer	5000 max vocab size

03 METHODS - setup

H2

REWARD MODEL TRAINING for a given utterance, it predicts whether it would likely have been followed by a CR by the caregiver. The goal is to provide caregiver-like reward to the input-based baseline's own produced utterances in the fine-tuning stage.

FINE-TUNING THROUGH RL using PPO for a maximum of 6000 steps and the best checkpoint is selected based on the mean reward.

For each fine-tuning step, they:

a) sample utterances from the language model (with the default temperature of 1)
b) compute the corresponding rewards based on the reward model
c) update the language model's weights using PPO

DETAILS OF THE STRATEGY ADOPTED
They use **rejection sampling** to discourage too-long and too-short utterances: –1 reward for < 3 tokens sentences or sentences that did not include the EOS token within 20 token

To obtain a diverse set of produced utterances: they randomly prompt the model with short beginnings (the first 1 to 2 tokens) of utterances from the pre-training data

Addition of an **entropy regularization term** (0.001) to the loss
Addition of a **small language modeling loss regularization term** (weighted by 0.001) to the loss and set the target KL-divergence to 2
All other **PPO hyperparameters are the same as implemented in the Huggingface** TRL library

03 METHODS - evaluation

To assess the effect of CR on grammar learning, they compare the language models' performance:

1. after language modeling pre-training
2. after fine-tuning

To **evaluate the grammaticality of produced utterances**, they sample 10K utterances from the model under evaluation and annotate them for grammaticality using two different models.

1. **Automatic classifier** (Nikolaus et al., 2024), specialized for Child–caregiver dialogue (handles colloquial & elliptical constructions)
   retrained without context for fair comparison

   Scoring scheme: Grammatical = 1
   Ambiguous = 0
   Ungrammatical = –1

2. Used an **off-the-shelf grammar correction model** (Rothe et al., 2021), trained on large-scale written English data
   not tailored for child-caregiver dialogue or spoken language, to which they feed the 10K utterances

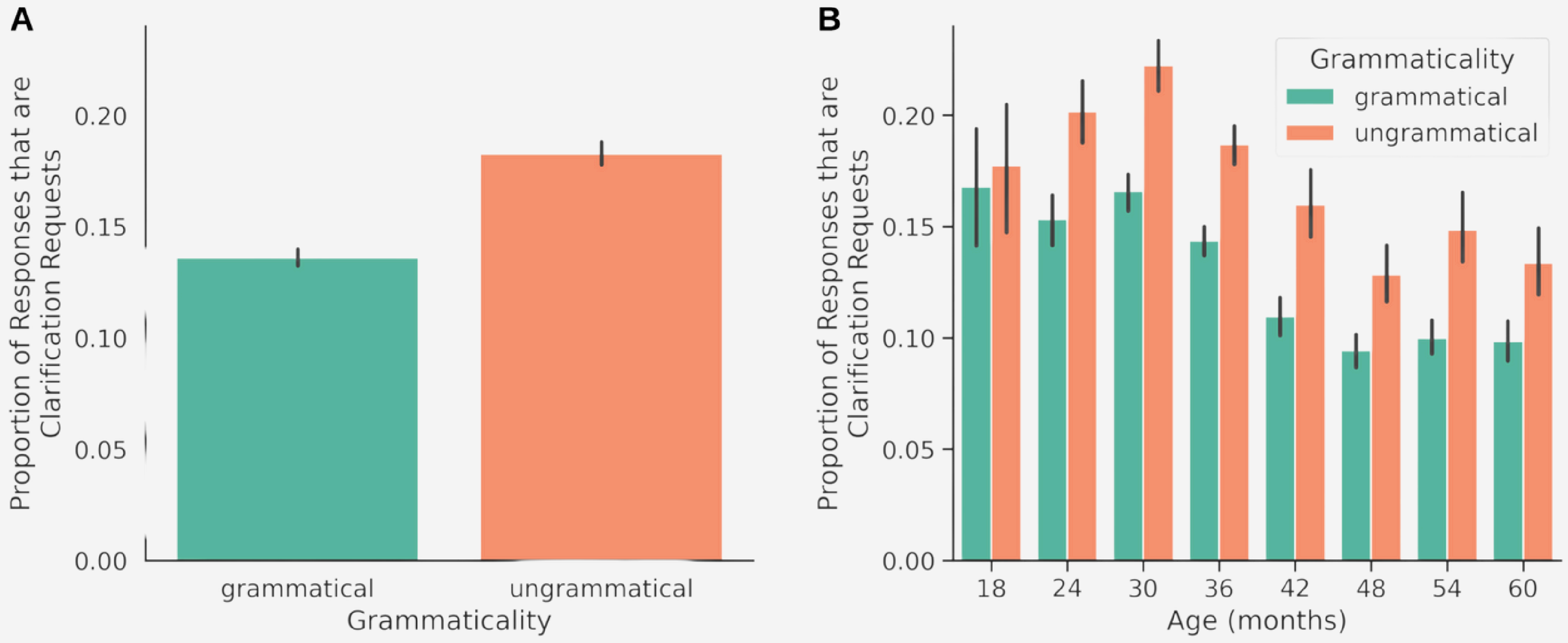
   Scoring scheme: if no correction was applied → Grammatical (score = 1)
   if correction occurred → Ungrammatical (score = 0)
   (punctuation/capitalization errors were ignored)

They evaluate the models' broad grammatical knowledge using **Zorro** (Huebner et al., 2021) and **Blimp** (Warstadt et al., 2020), filtering out the minimal pairs containing unseen words during the training.

04 RESULTS

CONTINGENCY OF CR

We find that CR is provided more frequently (than non-CR utterances) following a child's utterance that is *un*grammatical (mean: 0.183) when compared to the proportion of CR following a child's utterance that is grammatical (mean: 0.136). Figure B shows the same data, broken down by age group, and reveals a consistent effect across development.



They confirm this observation using a **Generalized Linear Model (GLM)**, predicting whether the caregiver responded with a CR as a function of the grammaticality of the child utterance:

$$response\_is\_clarification\_request \quad utterance\_grammaticality + (1|transcript\_id)$$

04 RESULTS

Baseline Models

Replicate prior findings: language modeling alone yields substantial syntactic knowledge

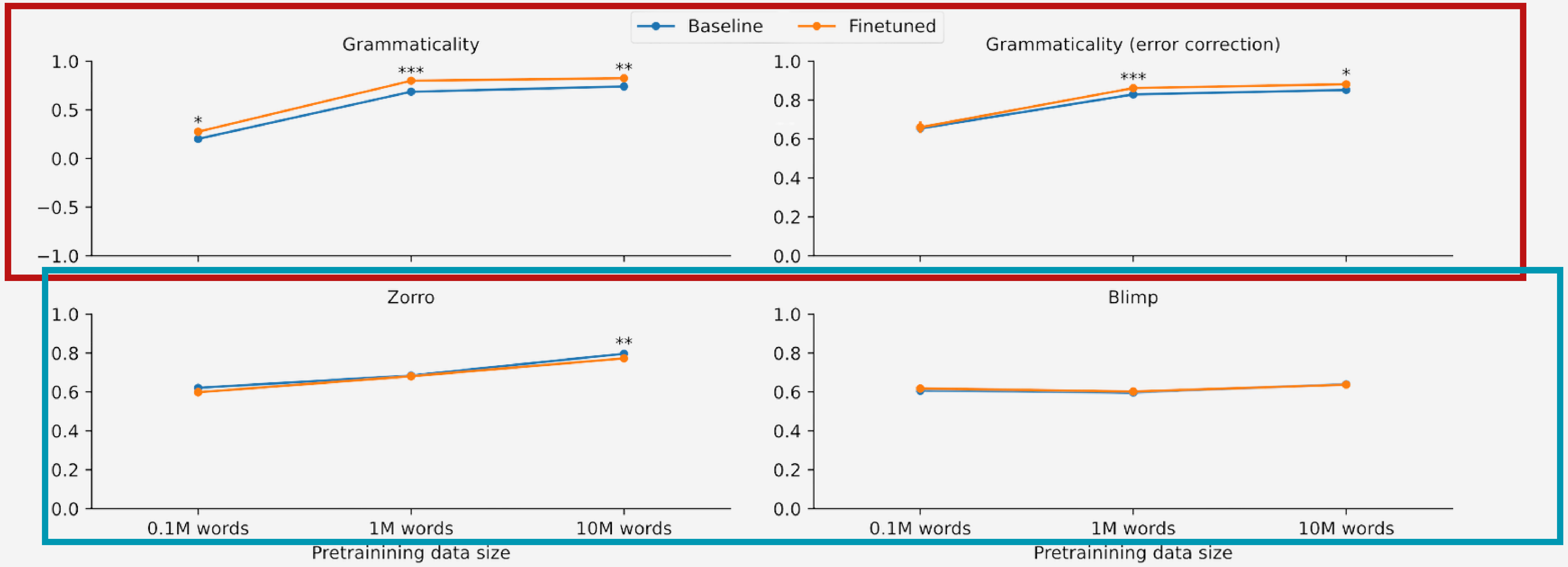
More input data (10M) → higher grammaticality, as models better approximate caregiver input

Reinforcement Learning (RL) Fine-Tuning

Significant improvement in grammaticality over baselines, observed across all data sizes (small, medium, large)

Improvements confirmed by grammar annotation and error correction models — except for the smallest data size (0.1M words)

For the **Grammar Evaluation Benchmarks** - no significant gains after fine-tuning.



04 RESULTS

Table 1

Samples of utterances produced by a baseline model (trained on 1M words) and the corresponding finetuned model along with grammaticality scores as evaluated with the two different models.

(a) *Utterances sampled from the baseline model*

Utterance	**Grammaticality**	**Grammaticality** (error correction)
it is not a little bit kind of a?	-1	0
there is the.	-1	1
yeah a little tiny hair i wonder if he is having a long.	-1	1
and if you want i tell you that.	-1	0
she is been in the museum for a tree.	-1	0
no not to get hurt.	-1	1
she is got some all clean.	-1	0
the end is there we go.	-1	0
oh, look at the.	-1	1
let's draw we have to buy her another picture if we can put them on.	-1	1
no it doesn't fit it?	0	1
okay i will get you some.	1	1
there is a baby in there.	1	1
where is the green book?	1	1
a pig.	1	1
and there is the fish.	1	1
she is coming home.	1	1
can you say please?	1	1
there is your eye.	1	1
what is that?	1	1
you don't think the doctor's on the phone.	1	1
let's see.	1	1
and a cow.	1	1
what happened?	1	1
oh here is one.	1	1

(b) *Utterances sampled from the finetuned model*

Utterance	**Grammaticality**	**Grammaticality** (error correction)
yeah a little tiny brown.	-1	1
she is got some all clean too right?	-1	0
what is this on?	-1	1
you want the, this one?	-1	0
that is right very great.	-1	0
i know what it is?	-1	1
oh, look at the.	-1	1
you don't really?	0	1
there is a baby.	1	1
no, you just want to read this one?	1	1
okay i will get the piece.	1	1
where is the green, is it a mommy?	1	1
a pig, yeah!	1	1
that is the green.	1	1
there is the car.	1	1
where are you going?	1	1
did you pick them up?	1	1
and what is that?	1	1
what is this one?	1	1
the end, there we go.	1	0
alright let's do this.	1	1
oh here is one.	1	1
there they are, they are right.	1	1
let's see.	1	1
no, that is a ball.	1	1
what are these honey?	1	1
where is the car?	1	1
what is it called?	1	1

05 DISCUSSION + follow up experiments

Clarification requests led to fewer ungrammatical utterances in the model

This is significant because:

- The feedback signal is vague (doesn't explain what error is done)
- It's noisy (CRs often follow grammatical sentences too)

Despite these limitations, the model still improved—supporting the idea that natural caregiver feedback can aid grammar learning

No improvement observed on NLP grammar benchmarks after fine-tuning

- Mismatch in error types: benchmarks may not cover the types of errors (e.g., omissions) that children and models most frequently make.
- Limited feedback coverage: Clarification requests (CRs) may not effectively target all relevant errors

Real grammar gains may not be captured by current benchmark tests

BEYOND CLARIFICATION REQUESTS
Clarification Requests (CRs) may miss certain benchmark-related errors, even if those errors occur in production

Follow-up experiment: Used an artificial reward model trained on benchmark data (Zorro & Blimp) to target these specific errors.
Same setup as before, but feedback was: b*inary* (still vague), but *less noisy* than CRs, since it directly focused on benchmark-relevant grammar errors.
Goal: Test whether more targeted feedback could yield better benchmark performance.

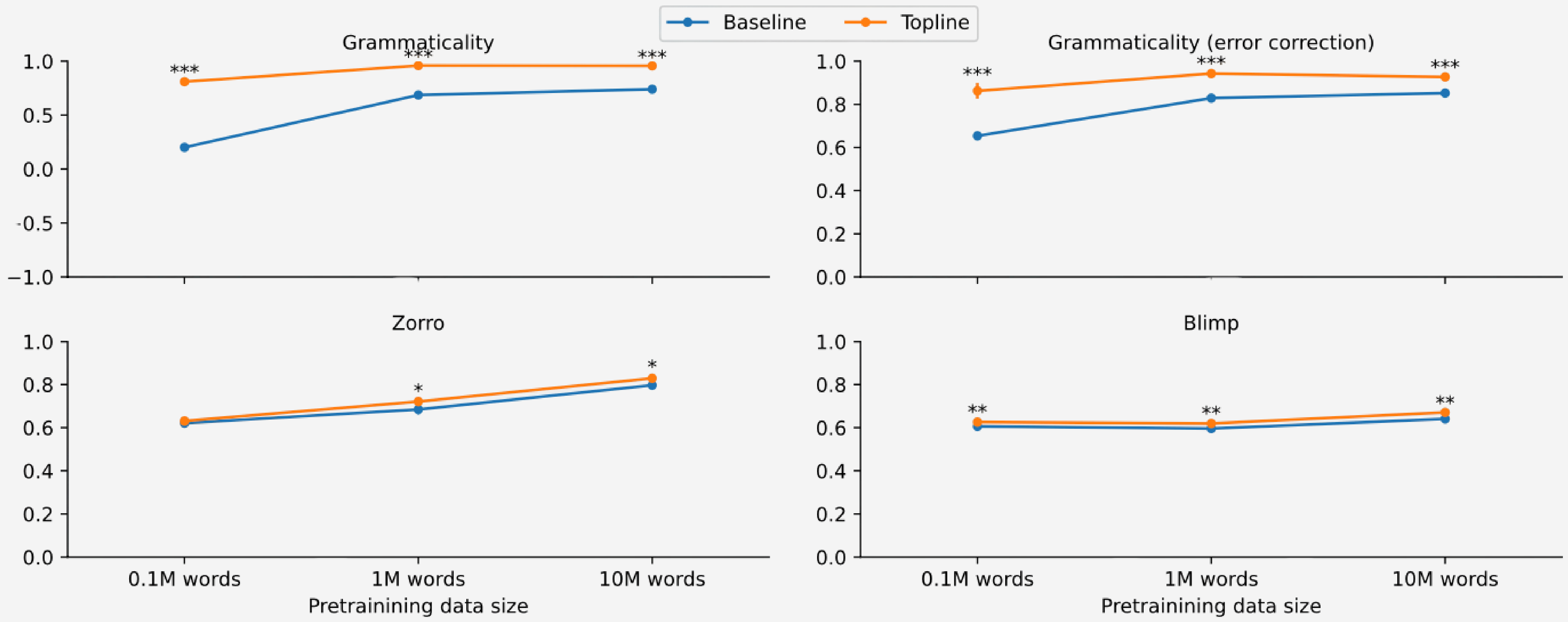
05 DISCUSSION + follow up experiments

Grammar scores improved significantly on Zorro & Blimp benchmarks **using artificial, targeted feedback**.

This shows some errors (e.g., subject-verb agreement, argument structures, irregular forms) are not well addressed by clarification requests (CRs) but can be learned with more precise feedback.

Suggests a ***"top-line"*** potential for grammar learning if stronger or multimodal caregiver feedback is used (e.g., gestures, intonation).

Opens the door for future work on combining feedback types for more effective language learning models.



05 What about caregivers' corrections?

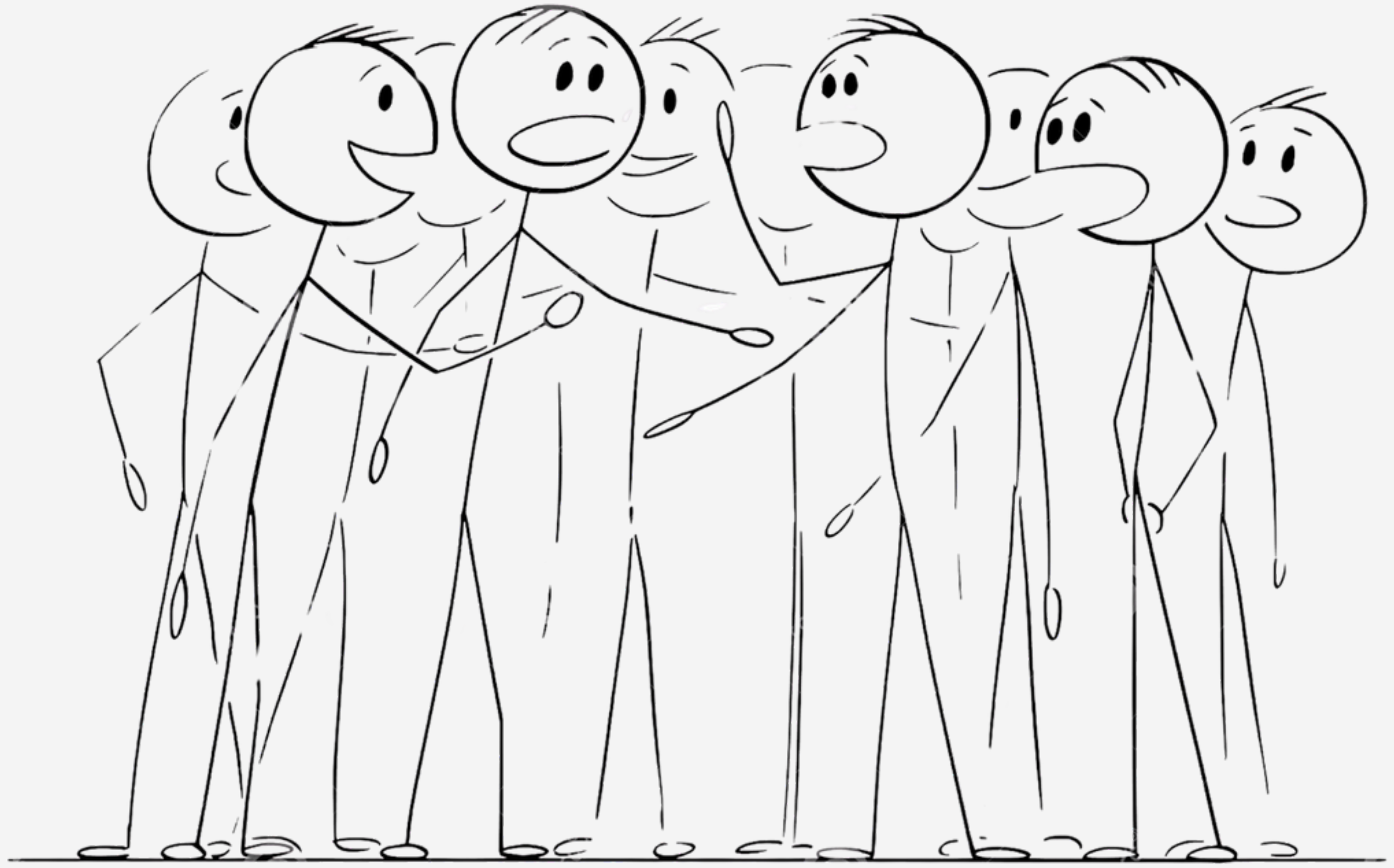
Contingent feedbacks > beyond negative feedbacks

One influential line of work focuses on explicit correction or reformulation by caregivers (Chouinard & Clark, 2003; Clark, 2020; Saxton, 1997)

Child: *It falled down!*
Caregiver: *It fell down?*

The caregiver's clarification request (a restricted offer) can be interpreted as **simultaneously providing negative evidence**—by implicitly rejecting the ungrammatical form ("falled")— and **contingent positive evidence**, by modeling the appropriate alternative ("fell").

This **dual function** has been proposed to support more effective learning than clarification requests that do not offer reformulation (e.g., "huh?")



TOWARDS DEVELOPMENTALLY PLAUSIBLE REWARDS: COMMUNICATIVE SUCCESS AS A LEARNING SIGNAL FOR INTERACTIVE LANGUAGE MODELS

Stöpler et al. 2025

(Heidelberg University, CNRS - Toulouse, ETH Zürich, University of California San Diego)

01 MOTIVATION

neural language models derive no utility from communication and unlike humans, their only objective is to predict the next token in a text authored by some other agent. This is a problem for at least 2 reasons:

1 **TRAINING OBJECTIVE**: LMs have the potential to transform computational modeling of human language processing and acquisition, but currently too divergent from humans to meet their full potential

2 **DATA EFFICIENCY**: humans are **more data-efficient** at acquiring language than LMs, and this may be due in part to the presence of an interactive learning signal.

Novel training regime for LMs that incorporates interactive learning. ***A speaker LM simulating a child****, learns from interacting with a* ***mature listener LM*** *and observing its degree of* ***communicative success****.*

01 SIGNAL FROM COMMUNICATIVE PRESSURE

1. **feasibility study**: showing that our notion of **communicative success** carries some **learning signal** for grammaticality. They find that communicative success degrades when the listener receives ungrammatical input.

2. conduct experiments **training T5** (Raffel et al., 2023) **from scratch** and in a **fine-tuning** setting

3. Systematically explore **ways to operationalize cost in production or comprehension** based on the *length or surprisal of the speaker's output*

4. Bottlenecks = constraints imposed on a language model's output (production) or input (comprehension), to **simulate cognitive pressures like processing cost**, to see how such constraints affect the model's linguistic behavior.

   4.1 **Length-based bottleneck**: the model is penalized for producing longer sentences.
   Encourages short, **"telegraphic" speech**—like in young children—often dropping function words (e.g., "is," "the," "of").

   4.2 **Surprisal-based bottleneck**: the model is penalized for producing high-surprisal words—i.e., words that are less predictable or expected given the context.
   Promotes outputs that are more predictable and natural-sounding and maintains grammatical properties similar to those of the unmodified LM.

01 THE ABSTRACT REFERENCE GAME

A **Reference Game** is a communicative task where one person (the speaker) must help another person (the listener) identify a specific referent, usually something like an object, image, or word—by providing a message that describes it (Lewis, 1969).

Language-only world: unlike traditional reference games that involve visual inputs (e.g., pictures), this version is purely text-based. This allows the model to handle:

- Abstract concepts
- Complex grammar
- Rich discourse structures

01 THE SUMMARIZATION GAME

The way in which we **operationalize** the abstract reference game uses a **combination of summarization** (El-Kassas et al., 2021) and **question–answering (QA)** (Khashabi et al., 2020) tasks common in natural language processing. The study models interaction as a 3-part communicative exchange:

FIRST

INFORMATION SHARING: the speaker sees a passage and must summarize it for a listener, balancing detail and brevity

SECOND

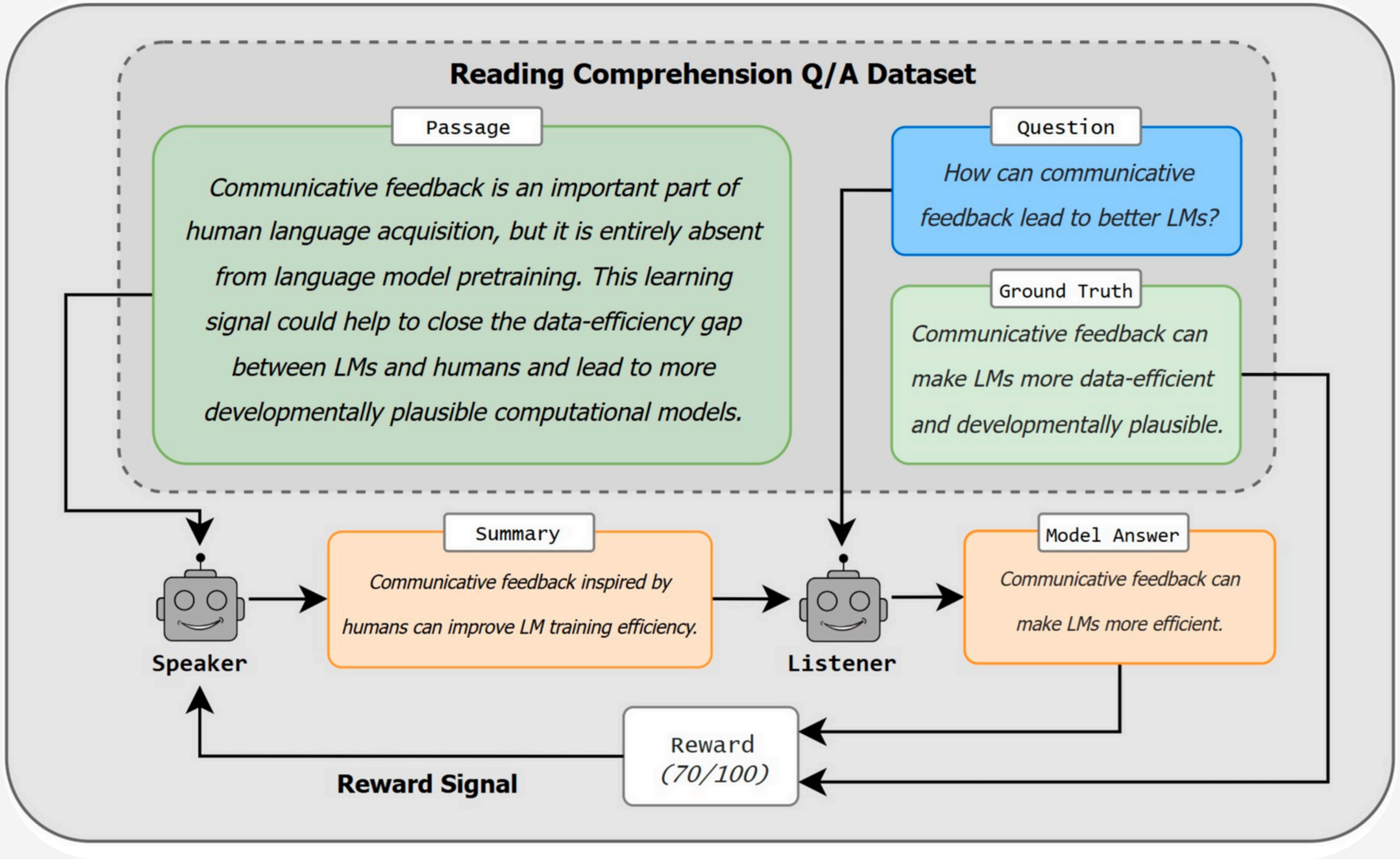
QUESTION RESOLUTION: the listener receives a question (based on the passage) and must answer it only using the speaker's summary

THIRD

FEEDBACKS AND EVALUATION: the listener's answer is compared to a ground-truth answer & the speaker is rewarded based on how accurate and complete the listener's answer is

They re-used already existing QA datasets to supply passages, questions, and answers.

01 THE SUMMARIZATION GAME



01 COMMUNICATION BOTTLENECK

Human communication is limited by production and comprehension **effort**.
Without constraints, the speaker could just repeat the full passage, bypassing true summarization, allowing the listener to receive maximal information, thus transmitting maximal information with no need for syntactic or semantic knowledge.

To model realistic communication, the study introduces two bottlenecks: **Number of Tokens** + **Surprisal**

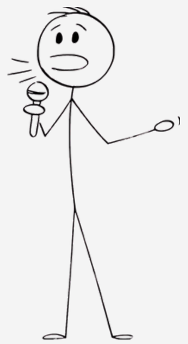
Cut-Off vs. Penalty

1. First, it could serve as a cut-off, **truncating the summary once the limit is reached**

2. Second, **it could be a penalty subtracted from the reward.**
   This is adopted as it yields a more continuous signal which could could lead to more stable training.

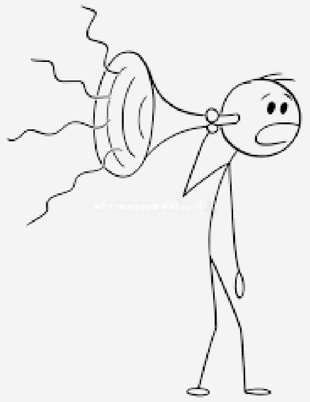
02 METHOD: SPEAKER AND LISTENER AGENTS

Simulates a language learning scenario:



Speaker = *learner* (updated during training)

- Trained from scratch on 70M-token C4 subset
- Fine-tuned version: pretrained T5 from Raffel et al. (2023) and fine-tuned on SQuAD 2.0 validation set



Listener = *expert* (frozen model)

- UnifiedQA (T5 fine-tuned on 17 QA datasets)

This has the added benefit of mitigating semantic drift (Lazaridou et al., 2020b), as the listener agent cannot adapt to innovations in the protocol introduced by the speaker.

Both roles use generative **LMs (T5)** to allow free-form summaries and answers.

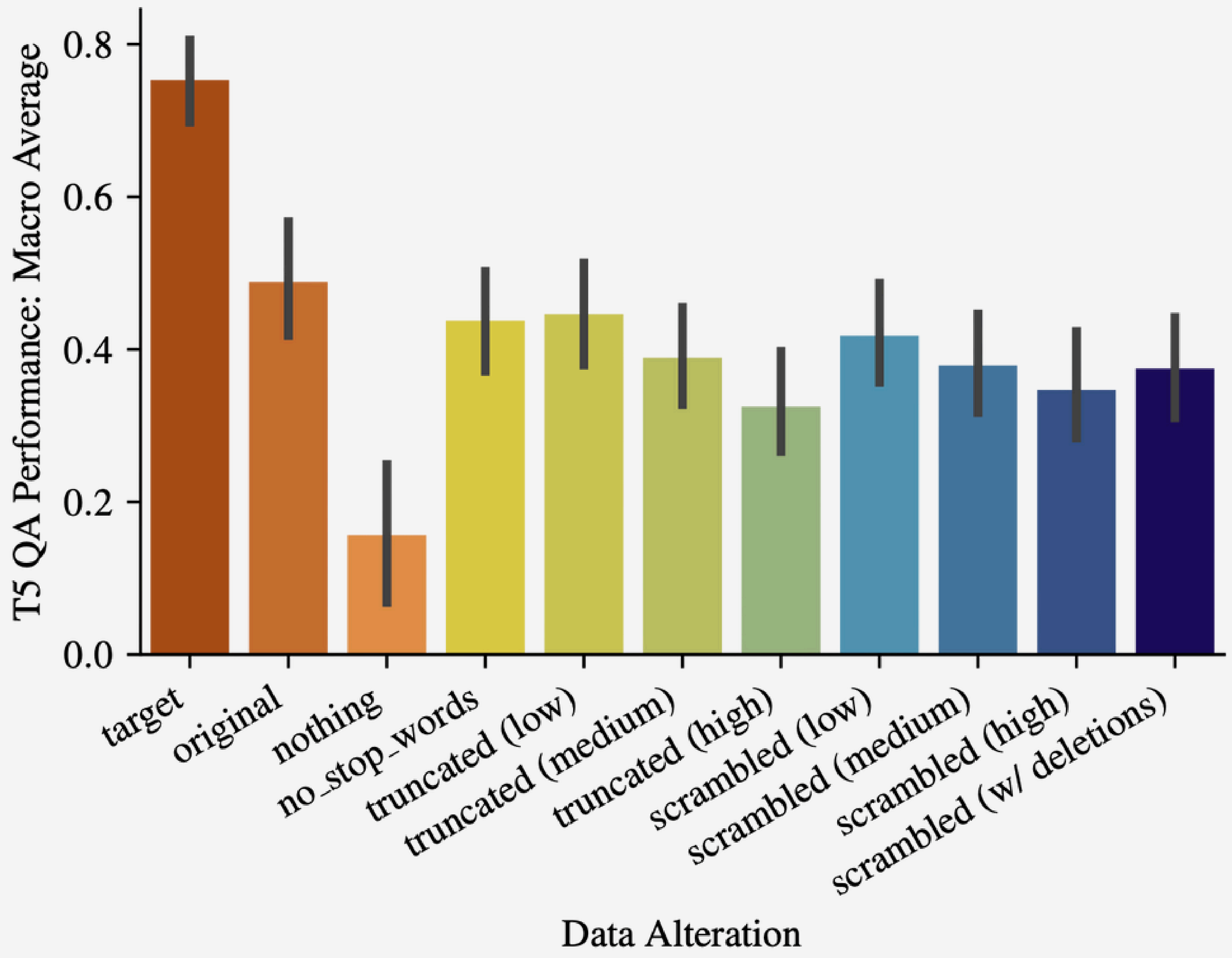
Methodological Details:

Reward: ROUGE-L F1 (handles short answers better than BLEU) balanced using hyperparameter $\lambda \in [0, 1]$

03 EXPERIMENT 1: FEASIBILITY STUDY

OBJECTIVE: provide a proof of concept that an abstract reference game provides a learning signal for language acquisition in LMs.

METHOD: measuring how the response of the listener model changes as a function of the quality of the passage it is provided.



03 EXPERIMENT 1: RESULTS

Skylines

- Best performance: when the listener sees the target answer
- Second Best: the full passage, then no context

Deletion Experiments

- *Removing stop words* → slight performance drop → shows incentive to maintain grammaticality
- *Random truncation* → listener performance drops proportionally to how much is omitted

Permutation Experiments

- *Word-order scrambling* (within sentences) harms syntax & semantics
- *Higher scrambling or deletion rates* → progressive performance degradation

03 EXPERIMENT 2: TRAIN FROM SCRATCH

Tested a randomly initialized language model to better simulate human language acquisition, beginning with next-word prediction training, then alternated with the summarization game objectives.

No bottleneck applied, to give the model maximal opportunity to learn.

RESULTS:

- Despite multiple runs**, no reward improvement was observed, models quickly degenerated into nonsensical output**
- **RL training from scratch is ineffective, even with supportive setup**

03

EXPERIMENT 3: FINE-TUNING

SETUP: fine-tuned a pretrained model (not from scratch), testing 5 levels of bottleneck strength ($\lambda \in \{0, 0.1, 0.5, 0.9, 1\}$) for both length and surprisal penalties.

KEY RESULTS:

- $\lambda = 0$ (no bottleneck): Reward improves, but outputs become longer, risk of drifting toward verbatim copying.
- Higher λ:
    - **Length penalty harms reward at $\lambda \geq 0.5$**.
    - **Surprisal penalty is more robust**, allowing better summaries even with high λ.

🗣️ **LANGUAGE DRIFT:**

- As reward ↑: summaries become closer to original, more verbose.
- As penalty ↑: telegraphic outputs emerge, especially with length bottleneck.

📉 **GRAMMATICALITY:**

- No improvement observed (via LanguageTool or BLiMP benchmarks).