Simulated Language Learning from Communicative Goals and Linguistic Input

Hao Zhu (zhuhao@cmu.edu)
Yonatan Bisk (ybisk@cs.cmu.edu)
Graham Neubig (gneubig@cs.cmu.edu)
Language Technologies Institute
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15217 USA

Abstract

Children do not learn language from passively analyzing correlations between language and observations, but from interaction with caregivers or peers. The non-nativist approach claims that the main driver of language learning should be to achieve communicative goals. Imitation, on the other hand, is another natural desire that many argue influences language learning. However, there are still gaps in the research on what roles communicative goals and imitating linguistic input play in language acquisition, due to the difficulty of performing comprehensive experiments with human learners. In this paper, we propose a computational framework using simulated experiments that allows us to compare the roles of the two drivers. Specifically, we simulate a two-way communication game between a speaker, corresponding to a language learner, and a listener, corresponding to a caregiver or teacher. The speaker's communicative goals are modeled as rewards for successful completion of a referential game, and imitation is performed by mimicking feedback from the listener. The listener adaptively chooses to give feedback and makes choices based on the speaker's utterances.

With empirical results on naturalistic visual and language data, we find that communicative goals play an important role in driving language learning, whereas imitation accelerates the learning process. We also find that (1) models trained with communicative goals tend to use minimal vocabulary and utterances and overextend them to concepts outside the original word meanings; (2) the strategy with which the listener provides feedback also influences the learning results and speed. Code and data for replicating the experiments are available[1] to spur future research on models for computational studies of language learning.

Keywords: Interaction; Language Learning; Referential Games; Reinforcement Learning; Communicative Goals; Linguistic Input

Introduction

Children learn a striking amount of language in their first few years of life – thousands of sounds, words, grammatical categories, and how to combine them into meaningful utterances. Unlike most recent machine learning models, which learn language from static existing text or images (?, ?), very young children do not learn language purely from observing visual – linguistic co-occurrences, e.g. watching television (?, ?, ?, ?, ?), but rather from interacting with their parents in conversations regarding family members, body parts, animals, foods and clothing, directed by the interest of the child (?, ?, ?). The challenge is then to understand how this learning process works and what internal and external factors influence it.

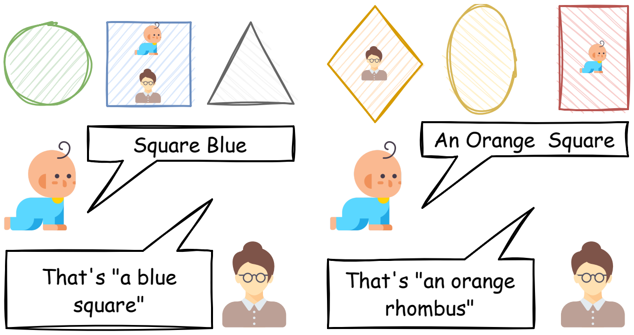

Figure 1: A child and adult playing shared-goal bidirectional communication games. The child learns from both communicative goals and the parent's feedback as linguistic input. On the left, the child uses incorrect word order to describe the shape in the middle, but the adult understands and gives corrective input. In contrast, on the right, the child uses "orange square" to describe the shape on the right, but the adult misinterprets and provides feedback for the shape on the left.

The most common and straightforward view is that children primarily use language as a tool to communicate (?, ?, ?). Just like learning to use other tools, one becomes more proficient via trial-and-error. Getting what they ask for, like asking for "applesauce" and receiving it instead of another object, reinforces the connection between entities and names. Conversely, failing to achieve a goal leads to a weaker connection or even negative reinforcement. Parents and adult members of the community also share intentions with children and respond to children's requests (?, ?), and thus children learn language from the use of language (?, ?). From this view, language is learned to convey meaning and reinforced by *communicative goals (CG)*, providing pressure to learn at least semantics, and perhaps also syntax to allow for disambiguation of more difficult concepts (?, ?)

Another way children learn language from their parents is through imitation, which has been studied for centuries since ? (1787). Although parents do not always explicitly point out grammatical mistakes in children's language, they offer corrective *linguistic input (LI)* to children based on their understanding of the meaning of the children's utterances (?, ?, ?). As a part of social learning, children imitate the feedback from their parents and learn the correlation between feedback

[1] `https://bit.ly/interactgym`

and the meaning they want to express. In this way, the fluency of a children's speech improves, but since the parents may not interpret the request correctly, the meaning of the feedback may align with the children's intent.

In this paper, we simulate this learning process in the context of *language games*, which have been the proving ground for various linguistics theories since their conception by ? (?).

Drawing an analogy to the child-parent conversation scenario, we model the child as a speaker agent, which generates utterances based on the target objects provided by the environment, and the parent as a listener agent, which responses to the child's utterances by choosing the objects and/or give corrective feedbacks. Based on this setup, we study the following research question

> *Can we form a computational speaker model that learns to speak from a skilled language listener under the communication game setting? What role do communicative goals and linguistic input play in this process?*

Our hypothesis is that communicative goals are the main driver of language learning in this formulation (**Communication Games** section), while linguistic input accelerates the learning process through syntax level supervision. To evaluate this, we use neural networks combined with heuristic rules to model the speakers and listeners (**Model** section). The learning process is implemented with a balance of reinforcement learning for CG and maximum likelihood estimation for LI (**Learning Process** section). We perform empirical experiments on communication games with MSCOCO (?, ?) images and captions (**Experiment** section). We find that CG contributes most to the game accuracy and also helps learning syntax as reflected by a fluency metric. We also find that different listener strategies also contribute to the success of language learning. Interesting, we also find that overextension in the resulting language of CG-driven models is very common, which is also common in early children speech (?, ?), while the same phenomenon does not often appear in models only trained with LI. Our results may provide evidence for usage-based language acquisition theory, the belief that language is acquired in the service of communicative functions (?, ?).

Communication Games

Following previous work on communicative agents learning to form communication pacts in referential games, we use asymmetric speaker-listener games (?, ?) with additional feedback channels.

A general goal-oriented language game provides an environment where the participants use language to communicate with each other to achieve the given goal. We consider the most basic setting of a collaborative referential game.

Procedure

As illustrated in Fig. **??**, in a communication game, the target image of the game $x \sim \mathcal{U}(I)$ is uniformly randomly sampled from the pool of images I, only visible to the speaker. N distractor images are randomly sampled from a distribution D_x^N. The target image and distractors are randomly shuffled before being shown to the listener to prevent any bias in the order. We denote the shuffled sequence of images as $\tilde{C}$, and order of the goal as i_g, i.e. $\tilde{C}_{i_g} = x$ and $\{\tilde{C}_i\}_{i \neq i_g}$ is a permutation of C.

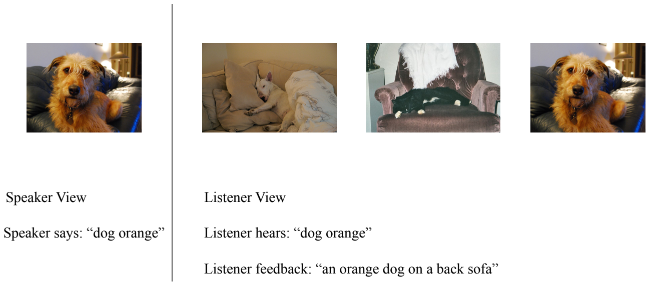
Speaker View
Speaker says: “dog orange”
Listener View
Listener hears: “dog orange”
Listener feedback: “an orange dog on a back sofa”

Figure 2: Game View. The speaker and listener have different knowledge about the game. Because the speaker does not know the distractor images (the white dog and black cat in this game), it needs to describe the image so that the listener could distinguish it from most distractors.

The speaker (modeling the child) takes the first turn in each game by describing the image in English. The listener (modeling the parent) then takes one of two actions based on the utterance u: (1) choose an image $\hat{i}$ or (2) perform no action $\hat{i} = \texttt{noop}$ (e.g. when they do not understand the utterance with enough confidence). Additionally, at the end of each game, the listener can choose to provide linguistic supervision to the speaker. At the end of each game, the speaker receives a reward based on the listener's action.

Reward

To model the communicative goals, we give positive rewards when the game is successful and negative rewards if the listener chooses the wrong image. In addition, we encourage the speaker to give unambiguous utterances by penalizing the noop action with a small negative reward $-1 < w_{\texttt{noop}} < 0$.[2]:

$$\mathcal{R}(i_g, \hat{i}) = \begin{cases} 1 & \hat{i} = i_g \\ w_{\texttt{noop}} & \hat{i} = \texttt{noop} \\ -1 & \text{otherwise} \end{cases} \qquad (1)$$

Speaker

As mentioned before, the participants consist of a speaker and a listener sending and receiving natural language messages. The speaker is a message-producing model defined by the *vocabulary* Σ; the space of *observations* I; and a *model* $f : I \to \Sigma^*$. The listener is an instruction-follower defined by the same vocabulary Σ, observation space I^{N+1}, and space of *actions* $[N+1]$ as the speaker; and a *model* $g : \Sigma^* \times I^{N+1} \to [N+1] \times \sigma^*$. Note that the listeners cannot

[2] We use a tighter lower bound of $w_{\texttt{noop}}$, so that a random choice is worse than no action: $w > \frac{1}{N} - 1$.

directly observe the goal, so the speakers need to use instructions to inform the listeners about the goal of each game.

Models

Speaker

The speaker is an image captioning model (?, ?, ?), which first encodes the goal image x with a pretrained ResNet (?, ?), and generates the utterance $u = u_i{}_{i=1}^{M}$ with an LSTM neural network (?, ?) in an auto-regressive fashion:

$$\begin{aligned}&\pi(u_i \mid u_1, u_2, \ldots, u_{i-1}, x)\\ &\propto \exp(w_{u_i}^T \mathtt{LSTM}(w_{u_1}, w_{u_2}, \ldots, w_{u_{i-1}}, h_0 = \mathtt{ResNet}(x))),\end{aligned} \tag{2}$$

where $w_{u_i} \in \mathbb{R}^{d_w}$ is the word embedding of word u_i.

Listener

A listener consists of two parts, a neural network-based ranker and a rule-based controller deciding whether to act and give language feedback to the speaker.

Given the utterance u, the listener ranks the images $\tilde{C}$ by the dot product between the LSTM embeddings of u and image embeddings encoded by the same pretrained ResNet as the speaker, i.e. for each image $\tilde{C}_i$, the score for ranking is

$$P_{\text{listener}}(i \mid x, C) \propto \exp(\mathtt{LSTM}^T(w_u)\mathtt{ResNet}(\tilde{C}_i)) \tag{3}$$

where w_u is a shorthand for the word embedding of all words in the sentence. Note that we use the same visual network for speakers and listeners, ignoring the differences between the visual perception of individuals, but the parameters language networks are not the same. We will discuss the method to acquire these parameters later in this section.

In human conversations, parents use a variety of techniques when giving feedback, including asking clarification questions and providing exemplar utterances. However, incorporating this open-ended feedback presents a huge challenge to the computational modeling of speakers. In this paper, we limit the feedback to full correct utterances for the goal image, which may be redundant or ineffective in many real world cases, but is general enough that most other kinds of feedback can be converted to it. We consider a listener controlled by both neural network rankers (as described above) and heuristic rules (Alg. **??**) which makes a choice when its confidence is high enough and gives feedback to the speaker if it thinks the utterance is not articulated well. Following (?, ?), we use the probability of prediction as the indicator of confidence. Alg. **??** has two parameters θ_1 and θ_2 which control making choices and giving feedback respectively. We will show the dramatic effects brought by these two parameters on language learning in the experiments. The golden utterances for images U^* are drawn from the captions provided in MSCOCO dataset (?, ?).

Listener Pretraining To model learning for a proficient language user, we need a good enough listener. Apart from θ_1 and θ_2 as well as the parameters in the ResNet, the parameters in the language network need pretraining. We use mini-batch stochastic gradient descent to optimize the following

Algorithm 1 Rule-Based Listener

```
Require: $\theta_1, \theta_2, P_{\text{listener}}, u, \tilde{C}, U^*$
                          ▷ $U^*$ is the golden description of images in $C$
  $g \leftarrow \arg\max_i P_{\text{listener}}(i \mid u, \tilde{C})$
  if $P_{\text{listener}}(g \mid u, \tilde{C}) \geq \theta_1$ then
      Make choice $f_{\text{listener}}(u, \tilde{C}) = g$
  else
      $f_{\text{listener}}(u, \tilde{C}) = \mathtt{noop}$
  end if
  if $P_{\text{listener}}(g \mid u, \tilde{C}) \leq \theta_2$ then
      Give feedback $h_{\text{listener}}(u, \tilde{C}) = U_g^*$
  end if    ▷ If the confidence is too low, the listener will not
  make a choice or give feedback.
```

$$\theta_{\text{listener}} = \arg\max_{\theta} \mathbb{E}_{x \sim \mathcal{U}(I), C \sim \mathcal{U}(I)^N} \log P_{\text{listener}}(i_g \mid x, C) \tag{4}$$

Learning Process

Objectives

The communicative goals and mimicking linguistic input can be modeled as two distinct learning objectives for the speaker network. Similar to children getting rewards from the environment if their request is fulfilled, and penalties otherwise, we use the expected game rewards as the objective for CG:

$$O_{CG} = \mathbb{E}_{x \sim \mathcal{U}(I), u \sim \pi(u|x), C \sim \mathcal{D}_x^N} \mathcal{R}(i_g, f_{\text{listener}}(u, \tilde{C})), \tag{5}$$

Note that the action space – the space of utterances – is discrete and non-differentiable, so we employ reinforcement learning to optimize the speaker policy π. Later in this section we give a brief introduction to PPO (?, ?), the RL method we used in the experiment.

Children's language models are reinforced if they recover the parent's corrective linguistic input. We thus model this objective as an maximum likelihood objective which measures parents' language in children's models.

$$O_{LI} = \mathbb{E}_{x \sim \mathcal{U}(I), u \sim \pi(u|x), C \sim \mathcal{D}_x^N} \log \pi(h_{\text{listener}}(u, \tilde{C}) \mid x). \tag{6}$$

This part is optimized with stochastic gradient descent.

To study the joint effect of both objectives, we adopt a multitask learning objective:

$$O_{\text{joint}} = \lambda O_{CG} + (1 - \lambda) O_{LI} \tag{7}$$

where λ is the coefficient balancing the two objectives and correlates with the importance of the CG objective.

Optimizing CG objective

Reinforcement learning methods are often employed to optimize non-differentiable objectives. In this subsection, we use the short hands state $s = \{u_i\}_{i=1}^{t-1}$, action $a = u_t$ at time step

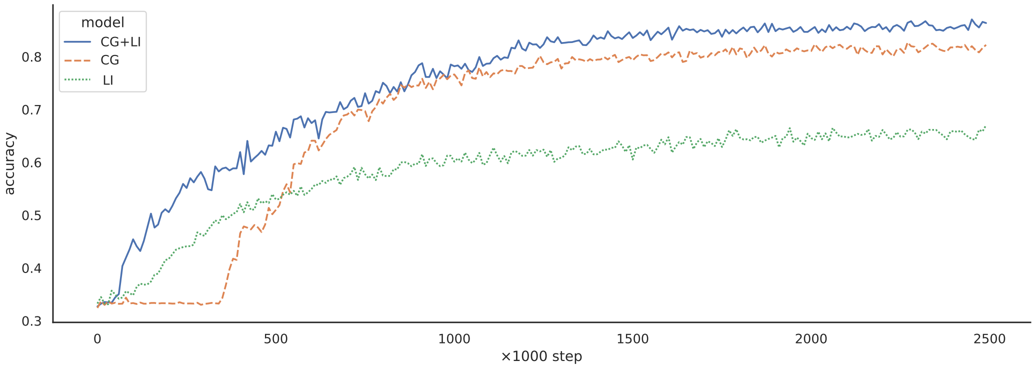

Figure 3: Accuracy change along training steps. We divide the training process into three stages. In Stage I (0-400k step), linguistic input leads to much steeper learning curve. In Stage II (400-1000k step) models with only linguistic input start to flatten out, but models driven by communicative goals continue to improve. And finally in Stage III (>1000k steps), models driven by communicative goals converge to a higher average reward than models with only linguistic input.

t for generating the utterance. The simplest RL method is policy gradient

$$L^{\text{PG}}(\pi) = \mathbb{E}_{s,a}(\pi(a \mid s)\mathcal{R}(s,a)). \tag{8}$$

As an on-policy method, PG optimizes its policy on the rollout data only once, which is inefficient. We can use importance sampling to reuse the data:

$$L^{\text{PG}}_{\pi_{\text{old}}}(\pi) = \mathbb{E}_{s,a}\left[r^{\pi_{\text{old}}}_{s,a}(\pi)A^{\pi_{\text{old}}}_{s,a}\right] + \eta(\pi_{\text{old}}) \tag{9}$$

where $r^{\pi_{\text{old}}}_{s,a}(\pi) = \pi(a|s)/\pi_{\text{old}}(a|s)$ is the *likelihood ratio* between the new policy π and the old policy π_{old} used to sample data, $A^{\pi_{\text{old}}}_{s,a} \triangleq \mathbb{E}[\mathcal{R}^{\gamma}_t|s_t = s, a_t = a; \pi_{\text{old}}] - \mathbb{E}[\mathcal{R}^{\gamma}_t|s_t = s; \pi_{\text{old}}]$ is the advantage value function of the old policy π_{old}. However, using this method does not guarantee a policy improvement. Therefore, TRPO (?, ?) and PPO (?, ?) are introduced with the basic idea of restricting the policy in a close distance from the old policy. PPO restricts the policy by a clipping function (?, ?)[3]

$$L^{\text{CLIP}}(\pi) = \mathbb{E}\left[\min\left(r_{s,a}(\pi)A_{s,a}, \mathcal{F}^{\text{CLIP}}\left(r_{s,a}(\pi), \varepsilon\right)A_{s,a}\right)\right] \tag{10}$$

where $\mathcal{F}^{\text{CLIP}}$ is defined as

$$\mathcal{F}^{\text{CLIP}}(r_{s,a}(\pi), \varepsilon) = \begin{cases} 1-\varepsilon & r_{s,a}(\pi) \leq 1-\varepsilon \\ 1+\varepsilon & r_{s,a}(\pi) \geq 1+\varepsilon \\ r_{s,a}(\pi) & \text{otherwise} \end{cases} \tag{11}$$

$(1-\varepsilon, 1+\varepsilon)$ is called the *clipping range*, and $0 < \varepsilon < 1$ is the parameter. Note that in theory most RL methods can be employed to optimize the CG objective. However, we use PPO here based on the trade-off of simplicity and relative good performance. We discuss other RL methods in the related works section.

Experiment

Game Setup

We use conventional split of MS COCO (?, ?). All of our neural networks are trained or pretrained on the training set, and all the results below are calculated on the test set.

In each game, after sampling the goal image x, the distractors are sampled from either uniform distribution $C \sim \mathcal{U}(I)^N$ (easy and default setting) or from a distribution skewed to the goal $C \sim \mathcal{D}_x^N$, where $\mathcal{D}_x(y) \propto e^{\|\boldsymbol{x}-\boldsymbol{y}\|_2}$ and $\boldsymbol{x}$ and $\boldsymbol{y}$ are embeddings from pretrained ResNet (?, ?) (hard setting).

Metrics

We use two metrics in the following experiments: (1) accuracy: the frequency of the listener choosing the goal among images; (2) fluency score, which reflects grammar quality of the sentence without considering semantics relatedness, following (?, ?)

$$\text{fluency} = \frac{1}{|u|}(\ln(p_M(u)) - \ln(p_U(u))) \tag{12}$$

We use GPT-2 large (?, ?) as p_M and a unigram model as p_U, both are fine-tuned/trained on MSCOCO.

What Drives Accuracy?

The first question we want to investigate is which signal is more important in learning semantically correct descriptions for the target image. In this paper, we use the listener's accuracy as a proxy to examine the semantic quality of generated descriptions. As shown in Fig. **??**, the accuracy of the LI-only model tops out at 60%, while models with the CG objective have significantly higher accuracy. However, the CG-only model needs about 400k steps to warm-up before dramatically improving on the similar performance of the combined model. With the help of LI, the CG+LI model (where $\lambda = 0.01$ is the best hyperparameter, used in all CG+LI models) not only has a faster improvement at the start of training,

[3] There are two variants of PPO: we refer to the one with clipping function as *PPO*, and refer to the one with adaptive KL penalty coefficient as *PPO-penalty* (?, ?).

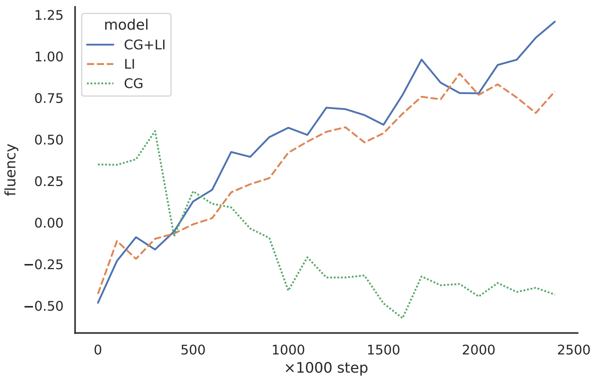

Figure 4: Fluency change along training steps. CG-only model decreases from 0.4 to -0.4, while CG+LI climbs from -0.5 to 1.25, and LI-only model climbs from -0.5 to 0.75.

but also achieves higher accuracy then CG-only model. From this result, we can see that CG is the main driver for conveying accurate information. The communication goal signal steers the model to output pragmatical descriptions that help the listener choose the correct target. In the hard setting, the CG+LI model and CG-only model both achieve 74% accuracy while the LI-only model only reaches 59%, which is a similar trend as the easy setting, thus confirming our conclusion still holds even if more detailed descriptions are needed for the game.

What Factors Help Fluency Learning?

The second question to investigate is which signal helps the speaker to learn to produce fluent language. Fig. **??** shows that LI is the main driver for learning to speak more fluently. The likely reason for the decreasing fluency of the CG-only model is the vocabulary shrinks and concentrates on a few words instead of all frequent ones in MSCOCO. In contrast, learning from linguistic inputs helps the model to fit the natural distribution of words. Later in this section, we will talk about the overextension of CG driven models. The improvement brought by LI may be the reason why CG+LI model does not need a warmup in Stage I in Fig. **??**.

Does the Listener's Strategy Affect to Learning?

In all previous experiments, we present results with $\theta_1 = 0.4$ and $\theta_2 = 0.9$ as the the thresholds for the listener strategy. In Fig. **??** we show the influence of these two parameters on the model's final accuracy. We find that the performance is very sensitive to the listener strategy. A small 0.05 change results in the difference between the best result and failure.

Overextension Phenomenon

Besides the experiments on CG and LI's influence on language learning, a signifcant difference between CG-driven models and LI-only model is overextension.

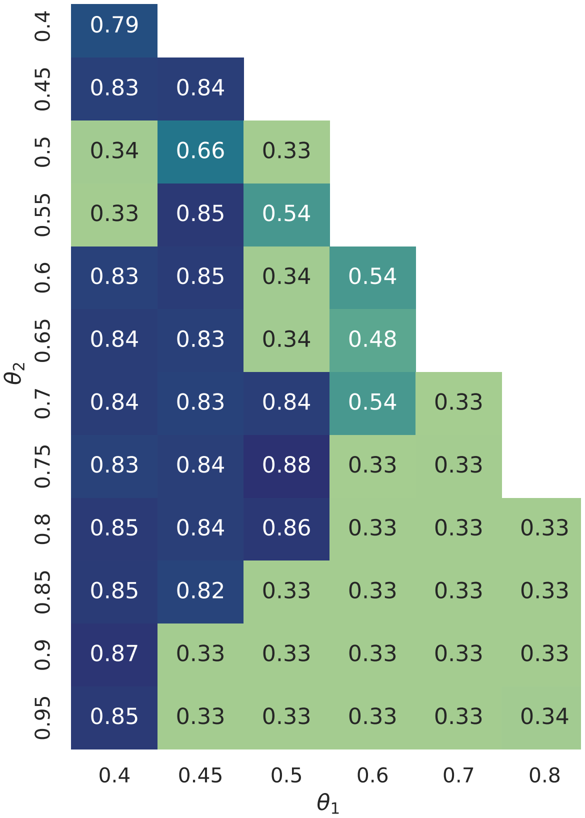

Figure 5: The influence of listener parameters to the accuracy of CG+LI model. Darker indicates better result, and 0.33 is the trivial accuracy since $N = 2$ in our experiments. Only showing the lower triangle, since $\theta_2 \geq \theta_1$ in Alg. **??**. The best parameters are the triangle region on the center left, and parameters outside it easily lead to failure.

To explore this, we randomly choose several nouns in the empirical vocabulary (words that exist in utterances) of CG+LI model. Most words exhibit intuitive cases of overextension. Some are based on color similarity, e.g., court; some are based on shape similarity, e.g. horse, giraffe, kite; while others are based on texture similarity, e.g. couch, pizza. We hypothesize there are a few possible reason for overextension in the model: (1) the shared visual perception – similar images to the speaker also look similar to the listener; (2) lack of linguistic inputs – with limited vocabulary, the RL model tests the acceptability of similar concepts; (3) the generality of the listener – the listener can understand the utterances just as we can interpret these errors. Although the models are making these errors, it may not necessarily be a bad thing. This phenomenon accounts for 40% of words used by 1;6 to 2;6 children (?, ?). This shows our formulation may be a good model of child language acquisition.

Related Work

Emergent Communication

Without natural language annotations, this pressure for the speaker and listener enables language emergence. (?, ?) first uses the same recurrent neural networks as the speaker and the listener to conduct emergent communication in referential game. Following their lead, (?, ?) study how emergent languages are grounded to the input images, and (?, ?) studies multi-turn communication via negotiation. (?, ?, ?) study the compositionally and systematicity of emergent languages. (?, ?) also explore the setting of training speaker with both reinforcement learning and MLE in referential games. To build a model that can communicate with humans, they start with a pretrained language model and use ground truth data of games in the experiment. Whereas we start from randomly initialized speakers and do not allow listeners' access to the goal to study language learning from scratch in the communication games.

Reinforcement Learning for Language Generation

Different reinforcement learning methods have been applied to language generation. On-policy methods include REINFORCE (?, ?, ?), actor-critic(?, ?, ?), policy gradient (?, ?, ?, ?), and off-policy methods include importance weighted policy gradient (?, ?, ?, ?), Q-learning (?, ?), and soft Q-learning (?, ?). In this paper, we use the most commonly used on-policy method PPO to optimize the CG objective. Experimenting with other methods is an interesting future direction.

Conclusion and Future Directions

In this paper, we propose a computational framework for language learning through communication games with both communicative goals and linguistic input objectives. We investigate the roles of CG and LI in general language learning in terms of conveying meaning and syntax learning. We also find that listener's strategy is important for language learning. This sheds light on child language learning – language usage may be the main driver, but without linguistic inputs language may be slow to acquire. Additionally, the adults' strategy in responding to the children's request is also important. Future work could further confirm this intuition by teaching human subjects (new) language with the best listener setting.

Acknowledgments

This work was supported by the National Science Foundation EAGER project (Award 2141751). We thank the anonymous reviewers and other lab members for their insightful comments and suggestions.

References

Anderson, D. R., Pempek, T. A. (2005). Television and very young children. *American behavioral scientist*, *48*(5), 505–522.

Bahdanau, D., Brakel, P., Xu, K., Goyal, A., Lowe, R., Pineau, J., … Bengio, Y. (2016). An actor-critic algorithm for sequence prediction. *arXiv preprint arXiv:1607.07086*.

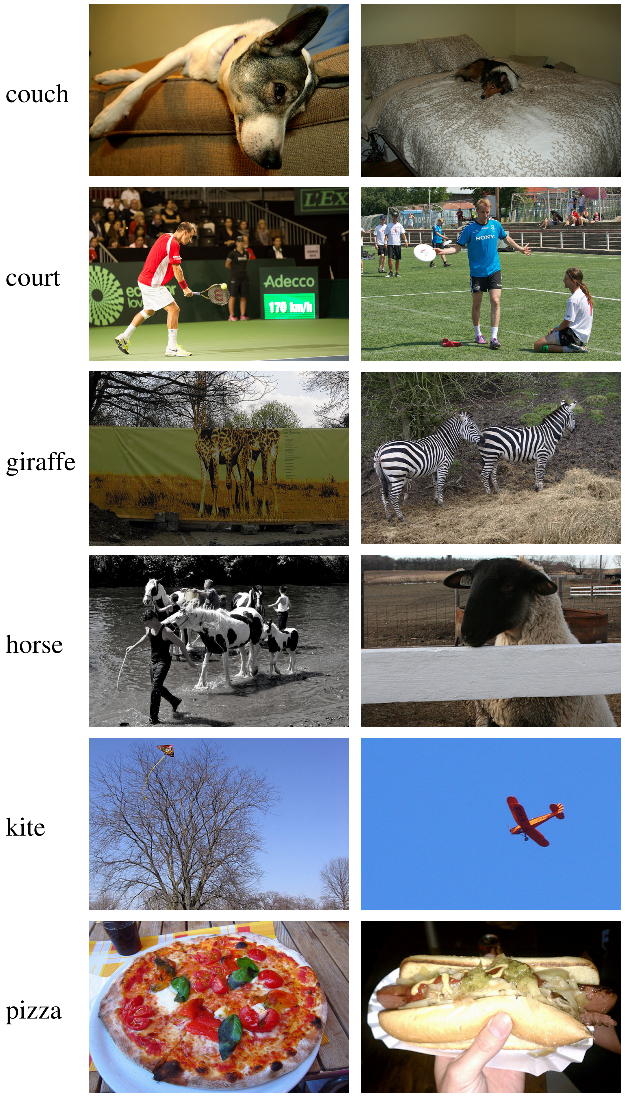

Figure 6: Overextension phenomenon visualized on part of the noun vocabulary learned by the CG+LI model. Images are cropped for visualization.

Batali, J. (1998). Computational simulations of the emergence of grammar. *Approach to the Evolution of Language*, 405–426.

Brown, R., Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard educational review*, *34*(2), 133–151.

Cao, K., Lazaridou, A., Lanctot, M., Leibo, J. Z., Tuyls, K., Clark, S. (2018). Emergent communication through negotiation. In *International conference on learning representations.*

Cazden, C. (1965). Environmental assistance to the child's acquisition of grammar. *unpublished Ph.D. thesis.*

Chaabouni, R., Kharitonov, E., Bouchacourt, D., Dupoux, E., Baroni, M. (2020). Compositionality and generalization in emergent languages. *arXiv preprint arXiv:2004.09124*.

Chen, Y.-C., Bansal, M. (2018). Fast abstractive summarization with reinforce-selected sentence rewriting. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 675–686).

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Farhadi, A., Hejrati, M., Sadeghi, M. A., Young, P., Rashtchian, C., Hockenmaier, J., Forsyth, D. (2010). Every picture tells a story: Generating sentences from images. In *European conference on computer vision* (pp. 15–29).

Ferguson, C. A. (1977). Baby talk as a simplified register.

Ferguson, C. A., DeBose, C. E. (1977). Simplified registers, broken language, and pidginization. *Pidgin and creole linguistics*, *99*, 125.

Guo, H., Tan, B., Liu, Z., Xing, E. P., Hu, Z. (2021). *Text generation with efficient (soft) q-learning.*

Gupta, A., Resnick, C., Foerster, J., Dai, A., Cho, K. (2020, July). Compositionality and capacity in emergent languages. In *Proceedings of the 5th workshop on representation learning for nlp* (pp. 34–38). Online: Association for Computational Linguistics.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Hendrycks, D., Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. *Proceedings of International Conference on Learning Representations*.

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Jaques, N., Shen, J. H., Ghandeharioun, A., Ferguson, C., Lapedriza, A., Jones, N., … Picard, R. (2020). Human-centric dialog training via offline reinforcement learning. *arXiv preprint arXiv:2010.05848.*

Kandasamy, K., Bachrach, Y., Tomioka, R., Tarlow, D., Carter, D. (2017). Batch policy gradient methods for improving neural conversation models. In *Iclr (poster).*

Kann, K., Rothe, S., Filippova, K. (2018). Sentence-level fluency evaluation: References help, but can be spared! In *Proceedings of the 22nd conference on computational natural language learning* (pp. 313–323).

Krcmar, M., Grela, B., Lin, K. (2007). Can toddlers learn vocabulary from television? an experimental approach. *Media Psychology*, *10*(1), 41–63.

Kuhl, P. K., Tsao, F.-M., Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.

Lazaridou, A., Peysakhovich, A., Baroni, M. (2016). Multi-agent cooperation and the emergence of (natural) language. In *International conference on learning representations.*

Lazaridou, A., Potapenko, A., Tieleman, O. (2020). Multi-agent communication meets natural language: Synergies between functional and structural language learning. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 7663–7674).

Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M., Gao, J. (2016). Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1192–1202).

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755).

Lowe, R., Gupta, A., Foerster, J., Kiela, D., Pineau, J. (2019). On the interaction between supervision and self-play in emergent communication. In *International conference on learning representations.*

MacWhinney, B., Bates, E. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, 3–73.

Mahowald, K., Diachek, E., Gibson, E., Fedorenko, E., Futrell, R. (2022). *Grammatical cues are largely, but not completely, redundant with word meanings in natural language.*

Mumme, D. L., Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child development*, *74*(1), 221–237.

Pang, R. Y., He, H. (2020). Text generation by learning from demonstrations. *arXiv preprint arXiv:2009.07839.*

Paulus, R., Xiong, C., Socher, R. (2017). A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304.*

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., … others (2021). Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020.*

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Ranzato, M., Chopra, S., Auli, M., Zaremba, W. (2015). Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732.*

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., Goel, V. (2017). Self-critical sequence training for image captioning. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7008–7024).

Rescorla, L. A. (1980). Overextension in early language development. *Journal of child language*, *7*(2), 321–335.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., Moritz, P. (2015). Trust region policy optimization. In *International conference on machine learning* (pp. 1889–1897).

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347.*

Tiedemann, D. (1787). *Disputatio de quaestione quae fuerit artium magicarum origo, quo modo illae ab asiae populis ad graecos atque romanos, et ab his ad ceteras gentes sint propagatae, etc*. Libr. Academica.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, *28*(5), 675–691.

Vinyals, O., Toshev, A., Bengio, S., Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3156–3164).

Wittgenstein, L. (1953). *Philosophical investigations*.

Wu, L., Tian, F., Qin, T., Lai, J., Liu, T.-Y. (2018). A study of reinforcement learning for neural machine translation. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3612–3621).

Zhou, L., Small, K., Rokhlenko, O., Elkan, C. (2017). End-to-end offline goal-oriented dialog policy learning via policy gradient. *arXiv preprint arXiv:1712.02838*.

Zhu, H., Neubig, G., Bisk, Y. (2021, 18–24 Jul). Few-shot language coordination by modeling theory of mind. In M. Meila T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 12901–12911). PMLR. Retrieved from `https://proceedings.mlr.press/v139/zhu21d.html`

Zipf, G. K. (1949). Human behavior and the principle of least effort: an introd. to human ecology.