Continual Adaptation for Efficient Machine Communication

Robert D. Hawkins[1], **Minae Kwon**[2], **Dorsa Sadigh**[2,3], **Noah D. Goodman**[1,2]

Departments of [1]Psychology, [2]Computer Science, and [3]Electrical Engineering

Stanford University, Stanford, CA 94305

{rxdh, mnkwon, ngoodman}@stanford.edu, dorsa@cs.stanford.edu

Abstract

To communicate with new partners in new contexts, humans rapidly form new linguistic conventions. Recent neural language models are able to comprehend and produce the existing conventions present in their training data, but are not able to flexibly and interactively adapt those conventions on the fly as humans do. We introduce an interactive repeated reference task as a benchmark for models of adaptation in communication and propose a regularized continual learning framework that allows an artificial agent initialized with a generic language model to more accurately and efficiently communicate with a partner over time. We evaluate this framework through simulations on COCO and in real-time reference game experiments with human partners.

1 Introduction

Communication depends on shared conventions about the meanings of words (Lewis, 1969), but the real-world demands of language use often require agents to go *beyond* fixed conventional meanings (Grice, 1975; Davidson, 1986). Recent work on *pragmatic* and *context-aware* models has approached this problem by equipping speaker and listener agents with the ability to explicitly reason about one another. Pragmatic reasoning allows listeners to infer richer intended meanings by considering counterfactual alternatives, and allows speakers to be appropriately informative, not merely truthful (Goodman and Frank, 2016; Andreas and Klein, 2016; Fried et al., 2018; Monroe et al., 2017; Vedantam et al., 2017).

These models have largely focused on one-shot settings, where the context is the immediate visual environment. In common interactive settings, however, the relevant context for pragmatic competence also includes the history of previous interactions with the same communication partner. Human in-
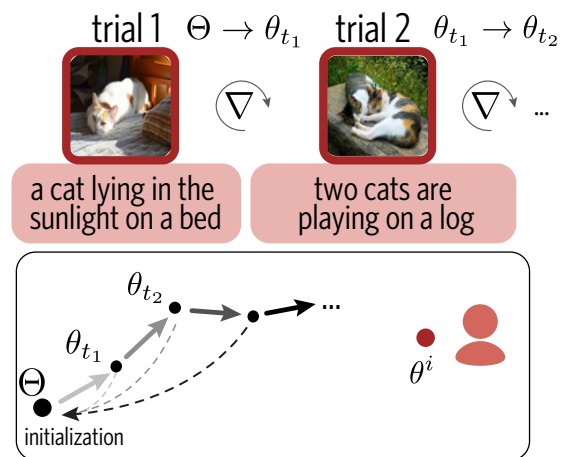

Figure 1: We introduce a regularized continual learning approach allowing agents initialized with a pretrained language model Θ to iteratively infer the language model θ^i used by a partner, over repeated interactions $\{t_1, t_2 \dots\}$ in an online reference game.

terlocutors are able to establish *ad hoc* conventions based on this history (Clark and Wilkes-Gibbs, 1986; Clark, 1996), allowing for increasingly *accurate* and *efficient* communication. Speakers can remain understandable while expending significantly fewer words (Krauss and Weinheimer, 1964; Orita et al., 2015; Staliūnaitė et al., 2018; Hawkins et al., 2020a; Stewart et al., 2020).

For example, consider a nurse visiting a bedridden patient at their home. The first time the patient asks the nurse to retrieve a particular medication, they must painstakingly identify a specific bottle, e.g. "the medicine for my back pain in a small blue medicine bottle labeled Flexeril in my bathroom." But after a week of care, they may just ask for the "back meds" and expect the nurse to know which bottle they mean. Such flexibility poses a challenge for current pragmatic models. For an artificial agent to establish new conventions, as humans do, it must go beyond pragmatic reasoning at the single-utterance timescale to *learn* about

its partners over longer timescales.

Here, we propose that the problem of *ad hoc* convention formation can be usefully re-formulated as an inference problem amenable to online domain adaptation. Our approach is motivated by a growing body of evidence in cognitive science that humans quickly re-calibrate their expectations about how language is used by different partners (Grodner and Sedivy, 2011; Yildirim et al., 2016). This empirical work highlights three key challenges facing a scalable adaptation approach. First, because the target data comes from intentional agents, *pragmatic reasoning* must be deployed throughout adaptation to strengthen inferences (Frank et al., 2009). Second, because the data is sparse, strong adaptation risks catastrophic forgetting; yet, human speakers are able to revert to their background expectations for the next interlocutor (Wilkes-Gibbs and Clark, 1992; Metzing and Brennan, 2003). Third, the ability to ground the meanings of later, shorter utterances (e.g. "back meds") in the use of earlier, longer utterances requires a *compositional* representation; otherwise the connection between the utterances is not clear (Hawkins et al., 2020a).

Our primary contribution is an online continual learning framework for transforming pragmatic agents into *adaptive* agents that can be deployed in real-time interactions. This framework is shown schematically in Fig. 1: after each trial, we take a small number of gradient steps to update beliefs about the language model used by the current partner. To evaluate our framework, we first introduce a benchmark *interactive* repeated reference task (Fig. 2) using contexts of natural images. In Sec. 3, we introduce the three core components of our algorithm: (i) a contrastive loss objective incorporating explicit pragmatic reasoning, (ii) a KL regularization objective to prevent overfitting or catastrophic forgetting, and (iii) a data augmentation step for compositionally assigning credit to sub-utterances. In Sec. 4, we report experiments demonstrating that this algorithm enables more effective communication with naive human partners over repeated interactions. Finally, in Sec. 5 we report a series of ablation studies showing that each component plays a necessary role, and close with a discussion of important areas for future research in Sec. 6

2 Related work

Personalizing language models. Adapting or personalizing language models is a classic problem of practical interest for NLP, where shifts in the data distribution are often found across test contexts (Kneser and Steinbiss, 1993; Riccardi and Gorin, 2000; Bellegarda, 2004; Ben-David et al., 2010). Our approach draws upon the idea of dynamically fine-tuning RNNs (Mikolov et al., 2010; Krause et al., 2017), which has successfully explained key patterns of human behavior in self-paced reading tasks (Van Schijndel and Linzen, 2018). We also draw on the regularization objectives proposed in this literatures (Li and Bilmes, 2007; Liu et al., 2016). However, the interactive communicative setting we consider poses several distinct challenges from traditional speech recognition (Miao and Metze, 2015) or text classification settings (Blitzer et al., 2007; Glorot et al., 2011) for which adaptation is typically considered. Partner-specific observations of language use are sparser, must be incorporated online, and are generated by intentional agents.

Incorporating discourse history. Previous work has incorporated discourse history in reference games using explicit co-reference detection (Roy et al., 2019) or contribution tracking (DeVault and Stone, 2009) techniques. An alternative approach is to include embeddings of the history as conditional input to the model at test time (Haber et al., 2019). Similar approaches have been proposed for sequential visual question answering (Ohsugi et al., 2019; Choi et al., 2018). Rather than pre-training a fixed, monolithic language model and incorporating shared history on top of this model at test time, we suggest that the underlying language model itself ought to be continually adapted over the course of an interaction.

Bayesian models of adaptation. Models of adaptation in cognitive science are typically formulated in terms of (hierarchical) Bayesian belief-updating based on evidence of language use (Kleinschmidt and Jaeger, 2015; Roettger and Franke, 2019; Delaney-Busch et al., 2019; Hawkins et al., 2017; Schuster and Degen, 2020). In these models, each new observation is taken as statistical evidence about the partner's language model, allowing pairs to coordinate on shared expectations and ground new conventions in their partner's previous behavior (see Sec. 3.2). While these models capture key theoretical properties of human adaptation, they do not scale well to natural-language applications, where neural networks are dominant.
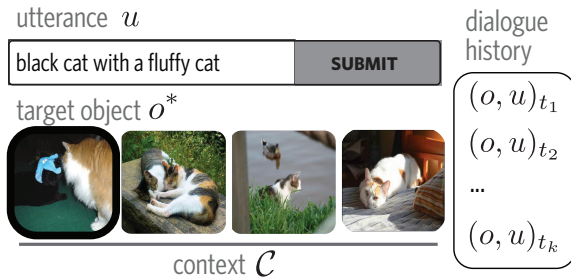
Figure 2: In a repeated reference game, a speaker agent must repeatedly communicate the identity of the same objects in context to a listener agent.

3 Approach

We begin by recasting convention formation as an online domain adaptation problem. As in previous computational approaches to pragmatics (e.g. Goodman and Frank, 2016; Andreas and Klein, 2016), we formulate this problem as an inference about another agent. The key theoretical idea is to expand the scope of pragmatic inference from the single-utterance timescale to evidence accumulated over longer timescales of an interaction. In addition to inferring a partner's intended meaning (or interpretation) for each individual utterance, an adaptive agent pools across previous utterances to infer the distinct but stable way their partner uses language. Under this inference framework, an agent must both (1) begin with background expectations about language shared across many partners, and (2) have a mechanism to rapidly learn the specific language model used by the current partner. Our work assumes a conventional neural language model as the starting point and focuses on the partner-specific inference problem. In this section, we describe our repeated reference game benchmark task (3.1), review the underlying problem as it has been previously formulated in a Bayesian framework (3.2), and finally describe our algorithm for adapting neural language models (3.3).

3.1 Repeated reference game task

As a benchmark for studying domain adaptation in communication, we use the *repeated reference game* task (Fig. 2), which has been widely used in cognitive science to study partner-specific adaptation in communication (Krauss and Weinheimer, 1964; Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992). In this task, a speaker agent and a listener agent are shown a context of images, $\mathcal{C}$ (e.g. four images of cats). On each trial, one of

these images is privately designated as the *target object*, o^*, for the speaker (e.g. the image with the thick border shown on the left). The speaker agent thus takes the pair $(o^*, \mathcal{C})$ as input and returns an utterance u (e.g. "black cat with a fluffy cat") that will allow the listener to select the target from $\mathcal{C}$. The listener agent takes $(u, \mathcal{C})$ as input and returns a softmax probability for each image, which it uses to make a selection. Both agents then receive feedback about the listener's selection and the identity of the target. Critically, the sequence of trials is constructed so that each image appears as the target several times. For example, our evaluations loop through each target six times, allowing us to observe how communication about each image changes as a function of dialogue history (see Fig. S1 in Supplementary Materials for examples).

3.2 The inference problem

We begin by assuming that agents represent the semantics of their language as a function relating natural language utterances u to actual states of the world o (here, images). We further assume that this function belongs to a family parameterized by θ, and denote the parameter used by a particular agent i with θ^i (see Fig. 1). If an artificial agent knows the true value of θ^i – their current partner's semantics[1] – they are in a better position to understand them, and to be understood in turn. However, because θ^i is not directly observable and θ varies across partners and contexts, it must be inferred (Bergen et al., 2016). Furthermore, it is in the agent's best interest to use its updated beliefs about its partner's θ to guide its own production and interpretation. An important consequence of this formulation is that conventionalization, the process by which parties converge on an efficient way to refer to something, emerges naturally as a consequence of mutual adaptation, the process by which each party independently tries to infer their interlocutor's language model (Smith et al., 2013; Hawkins et al., 2020b).

This is the central computational problem of adaptation, which we formalize as follows. Following Bayes Rule, the adaptive agent's beliefs about θ^i, conditioning on observations D^i from the shared history of interactions in that context, are:

$$P(\theta^i | D^i, \Theta) \propto P(D^i | \theta^i) P(\theta^i | \Theta) \qquad (1)$$

---

[1] Traditionally, this semantic function is truth-conditional, mapping utterance-state pairs to Boolean values, but recent approaches have shifted to more graded, real-valued functions such as those implemented by neural networks.

This formulation decomposes the inference into two terms, a prior term $P(\theta^i|\Theta)$ and a likelihood term $P(D^i|\theta^i)$.[2] The prior captures the idea that different partners share some general features of the semantics, represented by Θ, since they speak the same language; in the absence of partner-specific information, the agent ought to be regularized toward this background knowledge.

The likelihood term, on the other hand, accounts for direct evidence of language use. It represents an explicit forward model of an agent: different latent values of θ generate different observable actions. In other words, the standard single-utterance pragmatic inference problem is nested within the longer-timescale inference about θ. While explicit reasoning about the other agent is typically considered at the time of action selection (i.e. when the speaker is choosing an utterance, or when the listener is choosing a referent; Goodman and Frank, 2016; Andreas and Klein, 2016), this likelihood term importantly incorporates such reasoning at the time of *adaptation* (i.e. when updating beliefs about θ based on previous actions; Frank et al., 2009; Smith et al., 2013).

3.3 Continual adaptation for neural models

If we let θ be the weights of an image-captioning network, then the background knowledge shared across partners, Θ, corresponds to a pre-trained initialization, and conditioning on partner-specific data under a Bayesian prior corresponds to regularized gradient descent on θ. We exploit this connection to derive an online continual learning scheme that addresses the challenges of adapting to a human partner in a repeated reference game task.

Architecture and algorithm overview. Concretely, we consider an architecture that combines a convolutional visual encoder (ResNet-152) with an LSTM decoder (Vinyals et al., 2015). The LSTM takes a 300-dimensional embedding as input for each word in an utterance and its output is linearly projected back to a softmax distribution over the vocabulary size. To pass the visual feature vector computed by the encoder into the decoder, the final layer of ResNet was replaced by a fully-connected adapter layer. This layer was jointly pre-trained with the decoder on the COCO training corpus (Lin et al., 2014) and frozen. The COCO corpus con-

---

[2]For the rest of this paper, we only consider the case of adapting to one partner, so we will drop the partner index i.

Algorithm 1 Update step for adaptive model

1: **Input**: θ_t: weights at time t
2: **Output**: θ_{t+1}: updated weights
3: **Data**: (u_t, o_t): observed utterance and object
4: **for** step **do**
5:    sample augmented batch $\mathbf{u} \sim \mathcal{P}(u_t)$
6:    let $f_{\theta_t} = \log P_{\theta_t}(\mathbf{u}|o_t) + \log P_{\theta_t}(o_t|\mathbf{u}) - \mathrm{reg}(o_{1:t-1}, u_{1:t-1})$
7:    update $\theta_t \leftarrow \theta_t + \beta\nabla f_{\theta_t}$
8: **end for**

---

tains images of common objects, each annotated with multiple human captions. The CNN-LSTM architecture allows an agent to select utterances, by using beam search over captions given a target image as input, and also to select objects from the context, by evaluating the likelihood of the caption for each image in context and taking the most likely one.

Critically, we assume the agent will select actions on each trial using the value of θ it believes its partner to be using, so updating its own model is equivalent to updating expectations about its partner's model. Using the pre-trained model as our initialization, we can fine-tune the decoder weights (i.e. word embeddings, LSTM, and linear output layer) within a particular communicative interaction. Our algorithm is specified in Algorithm 1. Upon observing the utterance-object pair produced on each trial of the repeated reference game (Line 3), we take a small number of gradient steps updating the model weights to reflect the usage observed so far (Lines 4-7). Our adaptation objective function (Line 6) is built from combining a standard cross-entropy term with a KL-based regularization term to prevent catastrophic forgetting and a contrastive term to incorporate pragmatic reasoning about the visual context. In the following sections, we explain these terms and also introduce a final component of our approach: compositional data augmentation.

Utterance likelihood. For our benchmark repeated reference game, the data obtained on trial t is a paired observation of an utterance u and an intended object of reference o. The simplest learning objective for θ is the standard cross-entropy loss: the likelihood of this utterance being produced to convey the intended target in isolation: $P_\theta(u|o)$. This likelihood can be computed directly from the neural captioning model, where the probability of each word in $u = \{w_0, \ldots, w_\ell\}$ is given

by the softmax decoder output conditioned on the sentence so far, $P_{\theta_t}(w_i|o, w_{-i})$, so:

$$P_{\theta_t}(u|o) \propto \prod_{i<\ell} P_{\theta_t}(w_i|o, w_{-i}) \qquad (2)$$

Contrastive likelihood. The same object-utterance pairs can be viewed as being generated by a listener agent selecting *o relative to the other distractors* in the immediate context $\mathcal{C}$ of other objects. This reasoning requires inverting the captioning model to evaluate how well the utterance u describes each object in $\mathcal{C}$, and then normalizing:

$$P_{\theta_t}(o|u, \mathcal{C}, \theta_t) \propto P_{\theta_t}(u|o)P(o) \qquad (3)$$

This inversion is based on models of one-shot pragmatic inference in reference games (Goodman and Frank, 2016; Andreas and Klein, 2016; Vedantam et al., 2017; Cohn-Gordon et al., 2018). While optimizing the utterance likelihood serves to make the observed utterance more likely for the target in *isolation*, optimizing the contrastive likelihood allows the agent to make a stronger inference that it does *not* apply to the distractors.

KL Regularization. Fine-tuning repeatedly on a small number of data points presents a clear risk of catastrophic forgetting (Robins, 1995), losing our ability to produce or understand utterances for other images. While limiting the number of gradient steps keeps the adapted model somewhat close to the prior, we will show that this is not sufficient (see Sec. 5.1). Because small differences in weights can lead to large differences in behavior for neural models, we also consider a regularization that tethers the *behavior* of the adapted model close to the behavior at initialization. Specifically, we consider a *KL regularization* term that explicitly minimizes the divergence between the captioning model's output probabilities before and after fine-tuning for unseen images (Yu et al., 2013; Galashov et al., 2018). It is not tractable to take the KL divergence over the (nearly infinite) space of all possible natural-language utterances. Hence, we approximate the divergence incrementally by expanding from the maximum a posteriori (MAP) word denoted w^* at each step according to the initial model P_Θ (see Appendix A):

$$\sum_{i<\ell} D_{\mathrm{KL}}\left[P_\Theta(w_i|o, w^*_{-i}) \parallel P_{\theta_t}(w_i|o, w^*_{-i})\right] \quad (4)$$

where ℓ is the length of the MAP caption. This loss is then averaged across random images sampled from the full domain $\mathcal{O}$, not just those in context.

Compositional data augmentation. Agents should be able to infer previous successes on a longer utterance (e.g. "two men are sitting on a bench"), that the component parts of this utterance (e.g. "two men", "a bench") are also likely to convey the intended meaning. In the absence of a (weakly) compositional representation, a speaker has no way of doing credit assignment: observing that a listener successfully chose the target upon hearing a long utterance only provides further evidence for the full utterance. Fine-tuning an LSTM architecture will increase the likelihood of sub-strings to some extent after a successful selection, but this is insufficient for two reasons. First, not all sub-strings are syntactically well-formed referring expressions (e.g. "two men are"), and the LSTM lacks a syntactic representation to represent such coherence. Second, the likelihood of the full utterance will always be increased by more than any sub-utterance.

To address these problems, we explored a data augmentation step that introduces a stronger compositionality bias via referential entailments (Young et al., 2014). After each trial, we augmented the speaker's utterance u with a small denotation graph $D(u)$ containing the set of all noun phrases found in the syntactic dependency parse of u, and optimize our objective function on batches of these entailments. By independently updating expectations about well-formed entailments alongside the longer utterances that were actually produced, we hypothesized that our model could more naturally ground shorter, conventionalized labels in the shared history of successful understanding.

Local rehearsal. A second form of augmentation we explore is *local rehearsal*: at each step we include data from the history of interaction $D = \{(u, o)\}_{1:t}$ up to the current time t, to prevent overfitting to the most recent observation. In practice we subsample batches from the interaction history in a separate loss term with its own weighting coefficient, ensuring the new data point and a batch of its subphrase augmentations are used in every gradient step. We initialize D with the utterance the model generates for each object.

4   Interactive human evaluations

In this section, we evaluate our model's performance in *real-time interactions* with human speakers. Our artificial agent was paired with human partners to play a repeated reference game using
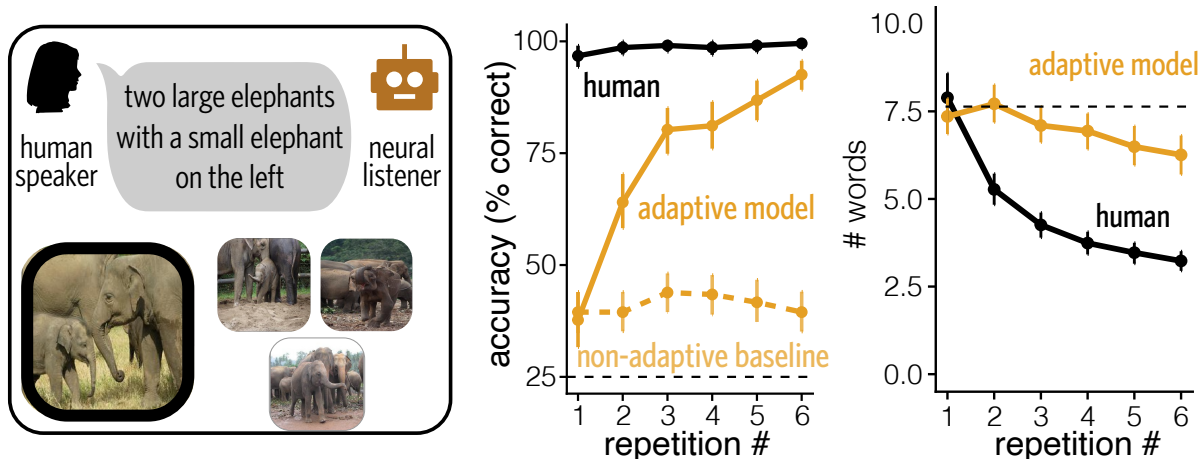
Figure 3: Communication becomes more efficient and accurate as our model adapts to a human speaker. Example contexts and utterances are shown. Error bars are bootstrapped 95% CIs.

images from the validation set of the COCO corpus (Lin et al., 2014; Chen et al., 2015) as the targets of reference. Critically, we constructed contexts to create a diagnostic mismatch between the COCO pre-training regime and the referential test regime. Specifically, we chose contexts such that the model's *accuracy* — the probability of identifying the target — would be poor at the outset.

To obtain appropriately challenging contexts, we used our pre-trained model's own visual encoder to find sets of highly similar images within the same category. We first extracted 256-dimensional feature vectors for each image from the final, fully-connected layer of the encoder. We then used these features to partition the images into 100 groups using a k-means algorithm, sampled one image from each cluster, and took its 3 nearest neighbors in feature space, yielding 100 unique contexts of 4 images each. This adversarial process explicitly identified contexts that our pre-trained captioning model would be poorly equipped to distinguish.

Human baselines. We first investigated the baseline performance of human speakers and listeners. We recruited 108 participants (54 pairs) from Amazon Mechanical Turk and automatically paired them into an interactive environment with a chatbox. For each pair, we sampled a context and constructed a sequence of 24 trials structured into 6 repetition blocks, where each of the 4 images appeared as the target once per block. We prevented the same target appearing twice in a row and scrambled the order of the images on each player's screen on each trial. We found that pairs of humans were highly accurate, with performance consistently near ceil-

ing (Fig. 3, black lines). At the same time, their utterances grew increasingly efficient: their utterances reduced in length across repeated interaction ($t = 25.8, p < 0.001$).[3]

4.1 Model performance

Next, we evaluated the performance of our adaptive model in the listener role (for a similar analysis of our model in the speaker role, see Appendix D). We recruited 57 additional participants from Amazon Mechanical Turk who were told they would be paired with an artificial agent learning how they talk. This task was identical to the one performed by pairs of humans, except we allowed only a single message to be sent through the chatbox on each trial. This message was sent to a server where the model weights from the previous trial were loaded to the GPU, used to generate a response, and updated for the next round. The approximate latency for the model to respond was 5-10s depending on how many games were running simultaneously.

For our adaptation objective function, we used a linear combination of the utterance and contrastive losses and the KL-regularization (see Appendix B for hyper-parameter settings). We also used local rehearsal and compositional data augmentation. While the pre-trained model initially performs much less accurately than humans, as expected, our adaptive listener shows rapid improvement in accuracy over the course of interaction (Fig. 3).

---

[3]Note that our contexts were selected to be challenging under the impoverished language prior of our pre-trained listener model, but were not expected to require any adaptation for human listeners to achieve high accuracy; see Hawkins et al. (2020a) for a more challenging stimulus domain used to elicit strong human adaptation.
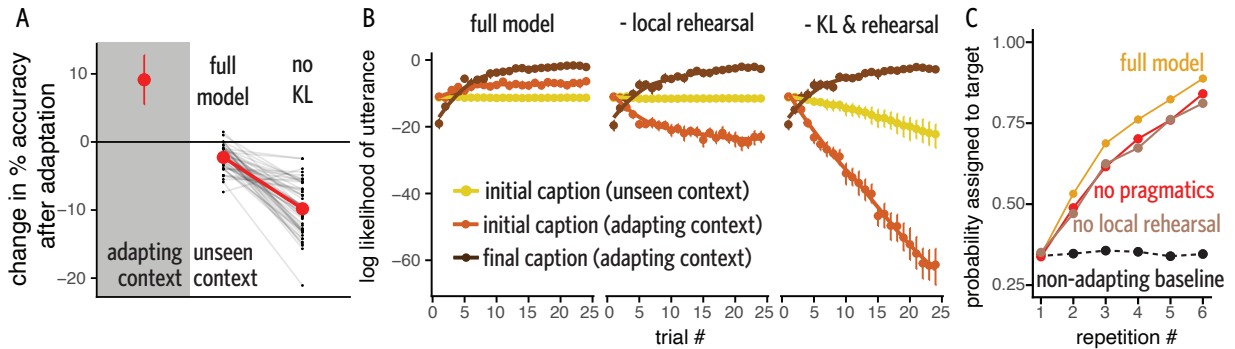
Figure 4: Ablation studies of the listener model. (A-B) KL regularization prevents catastrophic forgetting over the course of adaptation. (C) Local rehearsal and pragmatic reasoning independently contribute to successful listener adaptation. Error bars are bootstrapped 95% CIs.

In a mixed-effects logistic regression predicting trial-level accuracy, including pair- and image-level random effects, we found a significant increase in the probability of a correct response with successive repetitions, $z = 12.6$, $p < 0.001$, from 37% correct (slightly above chance levels of 25%) to 93% at the end. To test whether this success can be attributed to the initial quality of the listener model, or to *humans* adapting to a relatively unchanging model, we examined the performance of a non-adapting baseline (i.e. a model using the pre-trained model weights on every trial). We evaluated this baseline offline, using the utterances we recorded from the online games. This baseline showed no improvement, staying only slightly above chance accuracy over the course of the task.

5   Analysis

We now proceed to a series of ablation analyses that analyze the role played by each component of our approach. These analyses involve offline simulations conducted on the data we collected in the previous section.

5.1   KL regularization prevents catastrophic forgetting

We begin by testing the effectiveness of our KL regularization term (Eq. 4) for preventing catastrophic forgetting. We reasoned that changing expectations in the adaptation context should not interfere with expectations in other, unseen contexts. To directly analyze such interference, we adapted an ablated variant of our listener model over the course of a game with one context of images, and then measured its average accuracy identifying the target given the initial utterances produced by different speakers on different (unseen) contexts. We then

compared this test accuracy with the baseline accuracy achieved by an *unadapted* listener model.

We cross-validated these estimates over many adaptation contexts. Specifically, because the baseline was already close to chance on 'challenging' contexts (Fig. 3), we used an additional set of 52 human-human interactions we collected in easier contexts (where images belonged to different COCO categories) to better expose degradations in performance. While accuracy significantly increased compared to baseline in the adapting context for both variants, we found a 10% drop in accuracy on unseen contexts for the ablated variant with no KL term, compared to only a 2% drop in the model using the full loss ($t(51) = 12.2, p < 0.001$ in a paired t test; see Fig. 4A).

Next, to more thoroughly probe the *progression* of interference, we conducted a second analysis examining the likelihood assigned to different captions by the listener model over the course of adaptation. We tracked both the initial captions produced by the pre-trained initialization in the adapting context and in unseen contexts. To obtain unseen contexts, we sampled a set of images from COCO that were not used in our experiment, and generated a caption for each. We also generated initial captions for the *target* objects in the adapting context. We recorded the likelihood of all of these sampled captions under the model at the beginning and at each step of adaptation until the final round. Finally, we greedily generated an utterance for each target at the *end* and retrospectively evaluated its likelihood at earlier points during adaptation.

These three likelihood curves are compared for ablated models in Fig. 4B. By definition, the final caption in the adapting context becomes more likely in all cases (brown line). Without the lo-

414

cal rehearsal mechanism, the initial caption the model expected in the adapting context becomes less likely as it is replaced by the human partner's preferred caption (red line). Only when the KL term is removed, however, do we find interference with the model's expectations for unseen contexts (yellow line). Thus, we find that KL regularization plays a critical role in preventing catastrophic forgetting.

5.2 Pragmatics and local rehearsal improve listener performance

Next, we consider the contributions of other key components for success. Specifically, we constructed ablated variants of our model with no pragmatics (i.e. no contrastive loss term during adaption), and with no local rehearsal (i.e. no ability to keep training on batches from the history of the interaction). We simulated adaptation for these ablated variants on the 57 games where human speakers produced utterances for our listener model, and examined the probability assigned to the target after hearing each utterance (Fig. 4C). We found in a mixed-effects regression that each of these components independently contributes to success, as the ablated variants perform significantly worse than the full model ($z = 2.1$, $p = 0.03$ and $z = 3.6$, $p < 0.001$ for variants with no local rehearsal and no pragmatics, respectively; see Appendix C for regression details). Compared to an entirely non-adapting baseline, however, even these ablated variants improved over time.

6 Discussion

Relationship to human adaptation The theoretical ties between our approach and proposed cognitive models of human adaptation raises several questions. First, it is possible that improved performance could be driven by human *speakers* adapting in response to our listener agent's successes and errors rather than the other way around. While some degree of human adaptation is inevitable – for example, humans only seemed to shorten their utterances once our models' accuracy began to rise – human adaptation alone is insufficient to explain gains in accuracy. If these gains were due to human speakers gradually discovering utterances that a pre-trained (non-adapting) model could understand, we would expect some gains in the accuracy of our baseline non-adapting model over time. Furthermore, we found that the handful of human

speakers that dramatically changed their descriptions across rounds actually performed worse than those who adhered to consistent descriptions.

With this said, the extent of adaptation in human-computer dialogue is known to be affected by human participants' expectations about the artificial agent (Branigan et al., 2011; Koulouri et al., 2016), potentially including expectations about whether it will be adaptive or not. Bi-directional adaptation effects may be more pronounced in other dialogue settings where the human and model both speak, giving the human an opportunity to re-use utterances produced by the model. It will be important for future work to evaluate non-adaptive baselines *online* rather than offline, as we did, in order to observe exactly how humans respond to, or compensate for, non-adaptive agents.

Second, it is natural to ask how our model would perform in the *speaker* role with a human *listener*, using their (sparse) response success as feedback rather than their utterances. In ongoing work, we have found that the same approach allows a (pragmatic) model to converge to more efficient conventions in the speaker role (see Appendix D in supplemental), such that the same language model can flexibly switch between speaker and listener roles with the same human partner. Still, it is unlikely that this speaker model reduces in the same way as human speakers do (see Supplemental Fig. S1 for examples). Differences may reflect additional accessibility, grammaticality, or compositionality biases in humans; direct comparisons remain an open question for cognitive science.

Third, scaling the principles of computational-level Bayesian cognitive models to neural networks capable of adapting to natural language in practice required several algorithmic-level innovations which are not yet plausible proposals for human cognition (Marr, 2010). While our local rehearsal mechanism may be consistent with replay mechanisms in human memory, our KL regularization mechanism implausibly requires earlier parameter values of the model to be held in memory. Our data augmentation mechanism was introduced specifically to compensate for the inability of the LSTM architecture to propagate the use of a referring expression to its entailments, but we expect that human language processing mechanisms achieve this effect by different means. We expect further work to refine these algorithmic components as neural language models continue to advance.

Relationship to language learning Our work is also related to broader efforts to ground language learning and emergent communication in usage, where artificial agents are trained to use language *from scratch* by playing interactive reference games (Wang et al., 2016; Lazaridou et al., 2016; Wang et al., 2017; Chevalier-Boisvert et al., 2019). Rather than starting our agents from scratch, we have emphasized the need for continual, partner-specific learning even among mature language users with existing priors. This raises another question: how are these different timescales of learning related to one another? One possibility is that the need to quickly adapt one's language to new partners and contexts over short timescales may serve as a functional pressure shaping languages more broadly.

Recent theories in cognitive science have formalized this hypothesis in a *hierarchical* Bayesian model (Hawkins et al., 2020b). In this model, the prior Θ that an agent brings into subsequent interactions is updated to reflect the overall distribution of partner-specific models θ^i, thus balancing general and idiosyncratic language knowledge in a principled way. For neural language models, however, there is an apparent tension between the strong KL regularization required to *prevent* unwanted interference with background knowledge during partner-specific adaptation, leading to catastrophic forgetting, and the flexibility to generalize or transfer conventions to new communicative settings as required for language learning. We do not want to regularize so strongly that agents memorize conventions only applying to a single image that is completely reset after each interaction; instead, we wish to obtain a gradient of generalization across both referents and partners as a function of similarity (Markman and Makin, 1998).

One promising solution to this problem, motivated by connections between hierarchical Bayes and algorithms like MAML (Finn et al., 2017; Grant et al., 2018; Nagabandi et al., 2019), is to perform a meta-learning 'outer loop' updating the initialization Θ, taking into account the regularized, partner-specific 'inner loop' of adaptation for each θ^i. In principle, a meta-learning approach for neural language learning would distill abstract, shared aspects of language into a unified Θ, while still allowing for rapid *ad hoc* conventionalization. Still, cognitively plausible and scalable meta-learning algorithms remain an open area of research.

Limitations and future work While our evaluations were limited to a canonical CNN-RNN image captioning architecture, a key open question for future work is how our continual adaptation approach ought to be implemented for more complex, state-of-the-art architectures. One possibility, following the approach recently proposed by Jaech and Ostendorf (2018a), is to allow context (e.g. partner identity) to control a low-rank transformation of the weight matrix such that online fine-tuning can take place in a more compact context embedding space (Jaech and Ostendorf, 2018b).

Furthermore, while we adapted the entire parameterized RNN module end-to-end, future work should explore the effect of limiting adaption to subcomponents (e.g. only word embeddings) or expanding adaptation to additional model components such as attention weights or high-level visual representations. Beyond possible consequences for engineering better adaptive models, each of these variants corresponds to a distinct cognitive hypothesis about exactly *which* representations are being adapted on the fly in human communication.

A final area for future work is generalizing the forms of social feedback that can be used as data D^i for updating representations beyond the sparse choices in a reference game. In particular, forms of *repair* through question-asking or other non-referential dialogue acts may license stronger inferences about a partner's language model and allow misunderstandings to be resolved more quickly in challenging contexts (Drew, 1997; Dingemanse et al., 2015; Li et al., 2016). These forms of feedback may be particularly important for extending our approach beyond the benchmark task of repeated reference games to the more complex domains of real-world conversational tasks.

Conclusion Human language use is remarkably flexible, continuously adapting to the needs of the current situation. In this paper, we introduced a challenging repeated reference game benchmark for artificial agents, which requires such adaptability to succeed. We proposed a continual learning approach allowing agents to form context-specific conventions by fine-tuning general-purpose representations. Even when pre-trained models initially perform inaccurately or inefficiently, our approach allows such models to quickly adapt to their partner's language in the given context and thus become more accurate and more efficient using common ground.

Acknowledgments

All code and materials available at:
https://github.com/hawkrobe/
continual-adaptation

References

Jacob Andreas and Dan Klein. 2016. Reasoning about pragmatics with neural listeners and speakers. In *Proceedings of EMNLP*, pages 1173–1182.

Jerome R Bellegarda. 2004. Statistical language model adaptation: review and perspectives. *Speech Communication*, 42(1):93–108.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. 2010. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175.

Leon Bergen, Roger Levy, and Noah Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9(20).

John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of ACL*, pages 440–447.

Holly P Branigan, Martin J Pickering, Jamie Pearson, Janet F McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1):41–57.

Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C Lawrence Zitnick. 2015. Microsoft COCO captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*.

Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, and Yoshua Bengio. 2019. BabyAI: First steps towards grounded language learning with a human in the loop. In *Proceedings of the 7th International Conference on Learning Representations*.

Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wen-tau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. QUAC: Question answering in context. In *Proceedings of EMNLP*, pages 2174–2184.

Herbert H Clark. 1996. *Using language*. Cambridge University Press.

Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Reuben Cohn-Gordon, Noah Goodman, and Chris Potts. 2018. Pragmatically informative image captioning with character-level reference. In *Proceedings of NAACL*, pages 439–443.

Donald Davidson. 1986. A nice derangement of epitaphs. *Philosophical grounds of rationality: Intentions, categories, ends*, 4:157–174.

Nathaniel Delaney-Busch, Emily Morgan, Ellen Lau, and Gina R Kuperberg. 2019. Neural evidence for bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, 187:10–20.

David DeVault and Matthew Stone. 2009. Learning to interpret utterances using dialogue history. In *Proceedings of EACL*, pages 184–192.

Mark Dingemanse, Seán G Roberts, Julija Baranova, Joe Blythe, Paul Drew, Simeon Floyd, Rosa S Gisladottir, Kobin H Kendrick, Stephen C Levinson, Elizabeth Manrique, Giovanni Rossi, and N. J. Enfield. 2015. Universal principles in the repair of communication problems. *PloS one*, 10(9):e0136100.

Paul Drew. 1997. 'Open' class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1):69–101.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1126–1135.

Michael C Frank, Noah D Goodman, and Joshua B Tenenbaum. 2009. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5):578–585.

Daniel Fried, Jacob Andreas, and Dan Klein. 2018. Unified pragmatic models for generating and following instructions. In *Proceedings of NAACL*, pages 1951–1963.

Alexandre Galashov, Siddhant M Jayakumar, Leonard Hasenclever, Dhruva Tirumala, Jonathan Schwarz, Guillaume Desjardins, Wojciech M Czarnecki, Yee Whye Teh, Razvan Pascanu, and Nicolas Heess. 2018. Information asymmetry in KL-regularized RL. In *Proceedings of the 7th International Conference on Learning Representations*.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th International Conference on Machine Learning*, pages 513–520.

Noah D Goodman and Michael C Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818 – 829.

Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griffiths. 2018. Recasting gradient-based meta-learning as hierarchical bayes. In *Proceedings of the 6th International Conference on Learning Representations*.

H. P. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, pages 43–58. Academic Press, New York.

Daniel Grodner and Julie C Sedivy. 2011. The effect of speaker-specific information on pragmatic inferences. In *The processing and acquisition of reference*, volume 2327, pages 239–272. MIT Press.

Janosch Haber, Tim Baumgärtner, Ece Takmaz, Lieke Gelderloos, Elia Bruni, and Raquel Fernández. 2019. The PhotoBook dataset: Building common ground through visually-grounded dialogue. In *Proceedings of ACL*, pages 1895–1910.

Robert D Hawkins, Michael C Frank, and Noah D Goodman. 2017. Convention-formation in iterated reference games. In *Proceedings of the 39th annual meeting of the Cognitive Science Society*.

Robert D Hawkins, Michael C Frank, and Noah D Goodman. 2020a. Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, 44(6):e12845.

Robert D Hawkins, Noah D Goodman, Adele E Goldberg, and Thomas L Griffiths. 2020b. Generalizing meanings from partners to populations: Hierarchical inference supports convention formation on networks. In *Proceedings of the 42nd annual meeting of the Cognitive Science Society*.

Aaron Jaech and Mari Ostendorf. 2018a. Low-rank RNN adaptation for context-aware language modeling. *Transactions of the Association for Computational Linguistics*, 6:497–510.

Aaron Jaech and Mari Ostendorf. 2018b. Personalized language model for query auto-completion. In *Proceedings of ACL*, pages 700–705.

Dave F Kleinschmidt and T Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148.

Reinhard Kneser and Volker Steinbiss. 1993. On the dynamic adaptation of stochastic language models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 586–589.

Theodora Koulouri, Stanislao Lauria, and Robert D Macredie. 2016. Do (and say) as I say: Linguistic adaptation in human–computer dialogs. *Human–Computer Interaction*, 31(1):59–95.

Ben Krause, Emmanuel Kahembwe, Iain Murray, and Steve Renals. 2017. Dynamic evaluation of neural sequence models. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2771–2780.

Robert M Krauss and Sidney Weinheimer. 1964. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1(1-12):113–114.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2016. Multi-agent cooperation and the emergence of (natural) language. In *Proceedings of the 5th International Conference on Learning Representations*.

David Lewis. 1969. *Convention: A philosophical study*. Harvard University Press.

Jiwei Li, Alexander H Miller, Sumit Chopra, Marc'Aurelio Ranzato, and Jason Weston. 2016. Learning through dialogue interactions by asking questions. In *Proceedings of the 5th International Conference on Learning Representations*.

Xiao Li and Jeff Bilmes. 2007. A Bayesian divergence prior for classifier adaptation. In *Artificial Intelligence and Statistics*, pages 275–282.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft COCO: Common objects in context. In *Proceedings of ECCV*, pages 740–755. Springer.

Chaojun Liu, Yongqiang Wang, Kshitiz Kumar, and Yifan Gong. 2016. Investigations on speaker adaptation of LSTM RNN models for speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5020–5024.

Arthur B Markman and Valerie S Makin. 1998. Referential communication and category acquisition. *Journal of Experimental Psychology: General*, 127(4):331.

David Marr. 2010. *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Charles Metzing and Susan E Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213.

Yajie Miao and Florian Metze. 2015. On speaker adaptation of long short-term memory recurrent neural networks. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association*, pages 1101–1105.

Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Proceedings of the 11th annual conference of the International Speech Communication Association*, pages 1045–1048.

Will Monroe, Robert D. Hawkins, Noah D. Goodman, and Christopher Potts. 2017. Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338.

Anusha Nagabandi, Chelsea Finn, and Sergey Levine. 2019. Deep online learning via meta-learning: Continual adaptation for model-based RL. In *Proceedings of the 7th International Conference on Learning Representations*.

Yasuhito Ohsugi, Itsumi Saito, Kyosuke Nishida, Hisako Asano, and Junji Tomita. 2019. A simple but effective method to incorporate multi-turn context with BERT for conversational machine comprehension. In *Proceedings of the First Workshop on NLP for Conversational AI*, pages 11–17.

Naho Orita, Eliana Vornov, Naomi Feldman, and Hal Daumé III. 2015. Why discourse affects speakers' choice of referring expressions. In *Proceedings of ACL*, pages 1639–1649.

Giuseppe Riccardi and Allen L Gorin. 2000. Stochastic language adaptation over time and state in natural spoken dialog systems. *IEEE Transactions on Speech and Audio Processing*, 8(1):3–10.

Anthony Robins. 1995. Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2):123–146.

Timo B Roettger and Michael Franke. 2019. Evidential strength of intonational cues and rational adaptation to (un-) reliable intonation. *Cognitive Science*, 43(7):e12745.

Subhro Roy, Michael Noseworthy, Rohan Paul, Daehyung Park, and Nicholas Roy. 2019. Leveraging past references for robust language grounding. In *Proceedings of CoNLL*, pages 430–440.

Sebastian Schuster and Judith Degen. 2020. I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition*, 203:104285.

Nathaniel J Smith, Noah Goodman, and Michael Frank. 2013. Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in Neural Information Processing Systems*, pages 3039–3047.

Ieva Staliūnaitė, Hannah Rohde, Bonnie Webber, and Annie Louis. 2018. Getting to "Hearer-old": Charting referring expressions across time. In *Proceedings of EMNLP*, pages 4350–4359.

Ian Stewart, Diyi Yang, and Jacob Eisenstein. 2020. Characterizing collective attention via descriptor context: A case study of public discussions of crisis events. In *Proceedings of the 14th International AAAI Conference on Web and Social Media*, pages 650–660.

Marten Van Schijndel and Tal Linzen. 2018. A neural model of adaptation in reading. In *Proceedings of EMNLP*, pages 4704–4710.

Ramakrishna Vedantam, Samy Bengio, Kevin Murphy, Devi Parikh, and Gal Chechik. 2017. Context-aware captions from context-agnostic supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1070–1079.

Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164.

Sida I Wang, Samuel Ginn, Percy Liang, and Christoper D Manning. 2017. Naturalizing a programming language via interactive learning. In *Proceedings of ACL*, pages 929–938.

Sida I Wang, Percy Liang, and Christopher D Manning. 2016. Learning language games through interaction. In *Proceedings of ACL*, page 2368–2378.

Deanna Wilkes-Gibbs and Herbert H Clark. 1992. Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2):183–194.

Ilker Yildirim, Judith Degen, Michael K Tanenhaus, and T Florian Jaeger. 2016. Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, 87:128–143.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. 2014. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78.

Dong Yu, Kaisheng Yao, Hang Su, Gang Li, and Frank Seide. 2013. KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7893–7897.