Countering Language Drift via Visual Grounding

Jason Lee[†]**, Kyunghyun Cho**[†‡*]**, Douwe Kiela**[‡]

[†] New York University; [*] CIFAR Azrieli Global Scholar; [‡] Facebook AI Research
jasonlee@cs.nyu.edu, kyunghyun.cho@nyu.edu, dkiela@fb.com

Abstract

Emergent multi-agent communication protocols are very different from natural language and not easily interpretable by humans. We find that agents that were initially pretrained to produce natural language can also experience detrimental *language drift*: when a non-linguistic reward is used in a goal-based task, e.g. some scalar success metric, the communication protocol may easily and radically diverge from natural language. We recast translation as a multi-agent communication game and examine auxiliary training constraints for their effectiveness in mitigating language drift. We show that a combination of syntactic (language model likelihood) and semantic (visual grounding) constraints gives the best communication performance, allowing pre-trained agents to retain English syntax while learning to accurately convey the intended meaning.

1 Introduction

A long-standing goal of artificial intelligence research is to develop agents that can cooperate with other agents, including humans, to solve tasks. As Gauthier and Mordatch (2016) propose, one way to get closer to this goal is to develop agents that can flexibly use human language to coordinate with themselves and with humans.

Recently, there has been a renewed interest in multi-agent communication (Foerster et al., 2016; Lazaridou et al., 2016). While agents can be very effective in solving the tasks that they were trained on, their multi-agent communication protocols bear little resemblance to human languages. A major open question revolves around training multi-agent systems such that their communication protocols can be interpreted by humans.

One option is to pre-train in a supervised fashion with human language, but even then it is found that the protocols diverge quickly when the agents are

Intended message:	*2 elephants and 1 lion*
No constraints	*floopy globber*
Syntactic	*democracy is a political system*
Syntactic+Semantic	*a pair of elephants and a large feline*

Table 1: Examples of valid messages under different constraints. Without constraints, agents may invent an arbitrary protocol to communicate the intended message. Syntactic constraints enforce "Englishness" but not semantic correspondence. Semantic constraints, e.g. with visual grounding, can enforce the communication channel to retain the meaning of the message.

fine-tuned on an external reward, as Lewis et al. (2017) showed on a negotiation task. Indeed, language drift is to be expected if we are optimizing for an external non-linguistic reward, such as a reward based on whether or not two agents successfully accomplish a negotiation.

Language drift might be avoided by imposing a "naturalness" constraint, e.g. by factoring language model likelihood into the reward function. However, such a constraint only acts on the *syntax* of the generated language, ignoring its *semantics*. See Table 1 for an illustration of different constraints. As has been advocated by multi-modal semantics (Baroni, 2016; Kiela, 2017), we investigate if appropriate semantic constraints can be imposed on the generated language through (in this case visually) *grounding* its meaning in a different modality.

In order to carefully study this problem, we require a task where drift can be accurately measured. Inspired by Lee et al. (2018), we use a multi-modal machine translation (MMT) dataset (Multi30k; Elliott et al., 2016) to construct a new communication game: Two machine translation agents—i.e., encoder-decoder models with attention—are tasked with successfully translating source language se-

quences to the target language using a third pivot language as an intermediary. The first agent's decoder output is fed into the second agent's encoder as input. We employ policy gradient methods to train the first agent with the target language log-likelihood as reward. Thus, we effectively fine-tune two pre-trained machine translation agents via a pivot language, facilitating the study of its drift.

Contrary to alternative two-agent communication tasks such as navigation, game-playing or dialogue—which either don't have clearly defined metrics or easily available natural language data—this pivot-based translation allows us to check exactly whether the communicated sequence corresponds to the intended meaning, as well as to the gold standard sequence. In addition, every single utterance has very clear and well-known metrics such as BLEU and log-likelihood, allowing us to measure performance at every single step.

In what follows, we show that language drift happens, and quite dramatically so, when fine-tuning using policy gradients. Next, we investigate imposing syntactic conformity (i.e., "Englishness") via language model constraints, and show that this does somewhat mitigate drift, but does not lead to semantic correspondence. We then show that additionally imposing semantic constraints via (visual) grounding leads to the best retention of original syntax and intended semantics, and minimizes drift while improving performance. We conduct a token frequency analysis, which corroborates our hypothesis, and show that grounding causes the model to better preserve the token frequency distribution of the pivot language (English), while fine-tuning with language model constraints alone leads to a frequency distribution different from the original natural language.

The ability to keep drift in check opens up exciting possibilities for natural language processing research: we could maximize reward while *retaining* the "Englishness" of the decoder, with obvious benefits for interpretability and interaction with humans. One general use case would be fine-tuning a language model pre-trained on large amounts of data for a given generation task with limited data, which is especially interesting given the recent interest in pre-trained language models (Radford et al., 2019). For instance when training chit-chat dialogue agents, we often want to optimize for some very high-level reward, such as engagingness or consistency, with hardly enough data to learn

simple English grammar. The ability to fine-tune a pre-trained independent "language module", without drift, is an exciting prospect. With this work, we aim to take a step in that direction, and show that semantic constraints in the form of grounding play an important role.

2   Prior Work

Our work is inspired by recent work in protocols or languages that emerge from multi-agent interaction (Lazaridou et al., 2017; Lee et al., 2018; Andreas et al., 2017; Evtimova et al., 2018; Kottur et al., 2017; Havrylov and Titov, 2017; Mordatch and Abbeel, 2017). Work on the emergence of language in multi-agent settings goes back a long way (Steels, 1997; Nowak and Krakauer, 1999; Kirby, 2001; Briscoe, 2002; Skyrms, 2010). In our case, we are specifically interested in *tabula inscripta* agents that are already pre-trained to generate natural language, and we are primarily concerned with keeping their generated language as natural as possible during further training.

Reinforcement Learning has been applied to fine-tuning models for various natural language generation tasks, including summarization (Ranzato et al., 2015; Paulus et al., 2017), information retrieval (Nogueira and Cho, 2017), MT (Gu et al., 2017; Bahdanau et al., 2016) and dialogue (Li et al., 2017). Our work can be viewed as fine-tuning MT systems using an intermediary pivot language. In MT, there is a long line of work of pivot-based approaches, most notably Muraki (1986) and more recently with neural approaches (Wang et al., 2017; Cheng et al., 2017; Chen et al., 2018). There has also been work on using visual pivots directly (Hitschler et al., 2016; Nakayama and Nishida, 2017; Lee et al., 2018). Grounded language learning in general has been shown to give significant practical improvements in various natural language understanding tasks (Gella et al., 2017; Elliott and Kádár, 2017; Chrupała et al., 2015; Kiela et al., 2017; Kádár et al., 2018).

3   Task and Models

3.1   Communication Task

We recast pivot-based translation as a communication game involving two MT agents: Fr→En and En→De. See Figure 1. Our dataset consists of N triples of aligned sentences $\{\text{Fr}_k, \text{En}_k, \text{De}_k\}_{k=1}^{N}$. Note that En_k is only used for evaluation and is not required for training. We first feed the
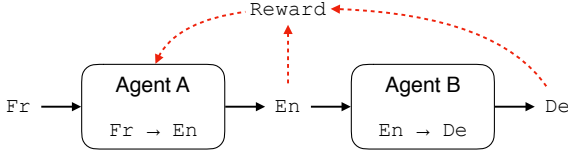
Figure 1: Diagram of our communication game.

French sentence Fr_k to Agent A, which generates an English message $\overline{En_k}$ as output. Agent B is then trained to maximize the log-likelihood of the ground truth German sentence given the English message, i.e. $\log p(De_k|\overline{En_k})$. Agent A is trained using REINFORCE (Williams, 1992) with reward $R = \log p_B(De_k|\overline{En_k})$.[1] This encourages Agent A to develop helpful communication policies for Agent B, and allows Agent B to adapt to Agent A's new policies. In other words: communication via the pivot language (English) is a success if we are able to translate the intended source sequence (French) into the desired target sequence (German).

Both agents are pre-trained on their respective tasks before communication, which means that the intermediate language starts off as English in the early stages of the communication game, where the goal is to translate French to German. This work examines what happens to the intermediate language as we fine-tune the system jointly for the given goal: will the agents keep communicating in English, or diverge? And if so, what can we do to prevent that from happening?

This particular task and setup directly addresses the problem of language drift, as the availability of ground truth references and well-understood metrics (e.g. BLEU) allows us to exactly measure the degree of language drift over time. The Fr→En→De BLEU score informs communication success, while (the relative change in) the Fr→En BLEU score captures the degree of language drift.

3.2 Constraints via Auxiliary Tasks

The action space of Agent A is $|V|^L$, where $|V|$ is the size of the vocabulary (approximately 20k) and L is the sequence length. We explore the two aforementioned constraints: a syntactic constraint via language modeling (LM) and a semantic constraint via grounding (G).

Language Model (LM) Given a language model pre-trained on a standard English corpus, the

(sentence-level) log-likelihood of the English message informs its general "Englishness". We incorporate this into the reward for Agent A, so that it learns to send messages that are plausible English.[2] Reward for Agent A is:

$$R_k^{\text{LM}} = \log p_B(De_k|\overline{En_k}) + \beta_{LM} \, \log p_{LM}(\overline{En_k}).$$

Grounding Model (G) Let us assume we have access to a set of images $\{Img_k\}$ associated with each triple $\{Fr_k, En_k, De_k\}$. Given a pre-trained image-caption retrieval model, such as VSE++ (Faghri et al., 2018), the log-likelihood of the image given the English message (and vice versa) captures how much the English message is grounded in the original semantic content (Kiela et al., 2017). We incorporate the ranking loss into Agent A's reward.

$$R_k^{\text{G}} = \log p_B(De_k|\overline{En_k}) + \beta_G \, \log p_G(Img_k|\overline{En_k}).$$

β_{LM} and β_G are hyperparameters.

3.3 Training Objective

Let us denote the t-th token in the k-th English training example with En_k^t, the actual reward and the state-dependent baseline in the k-th training example as R_k and $\overline{R_k^t}$.

Policy Gradient Training At decoding timestep t, Agent A takes an action (outputs token $\overline{En_k^t}$) given an environment (previous hidden states and previous token $\overline{En_k^{t-1}}$). It receives reward R_k at the end of the sequence, from which we subtract a state-dependent baseline $\overline{R_k^t}$ to reduce variance. Therefore, we maximize $(R_k - \overline{R_k^t}) \log p(\overline{En_k^t}|\overline{En_k^{<t}}, Fr_k)$. In addition, we employ entropy regularization on Agent A's decoder to encourage exploration. Hence, Agent A's overall objective function is given as:

$$\mathbb{L}_A = \sum_{k=1}^{N} \Bigg\{ \sum_{t=1}^{T_k} \Big\{ \alpha_{\text{entr}} \, H\big(p(\overline{En_k^t}|\overline{En_k^{<t}}, Fr_k)\big)$$
$$- \alpha_b \, \text{MSE}(R_k, \overline{R_k^t})$$
$$+ \alpha_{\text{pg}} \, (R_k - \overline{R_k^t}) \log p(\overline{En_k^t}|\overline{En_k^{<t}}, Fr_k) \Big\} \Bigg\},$$

---

[1] We use subscript B to denote that the probability is computed with Agent B.

[2] We also experimented with a dense LM reward on the word-level, but found this to lead to worse performance. We hypothesize that the model might be focusing too much on the dense LM reward, ignoring the sparse reward for the communication task and leading to poor performance. We did not use BLEU as it is a corpus-level metric.

where H and MSE denote entropy and mean squared error losses. T_k is the maximum decoding timestep in the k-th training example.

Cross Entropy Training Agent B is trained using standard cross entropy loss, i.e.

$$\mathbb{L}_B = \sum_{k=1}^{N} \Big\{ \sum_{t=1}^{T_k} \log p(\text{De}_k^t | \text{De}_k^{<t}, \overline{\text{En}_k}) \Big\}.$$

We jointly train both agents by maximizing $\mathbb{L} = \mathbb{L}_A + \mathbb{L}_B$.

4   Experimental Settings

In this section we provide the details of our experimental setup: a Fr→X→De translation task where the intermediate language X is initialized as English, and subsequently fine-tuned with policy gradient methods. The model is trained either with no constraints (**PG**), syntactic constraints via language modeling (**PG+LM**) or both syntactic and semantic constraints via language modeling and grounding (**PG+LM+G**).

Datasets Agents A and B are initially pre-trained on IWSLT Fr→En and En→De, respectively (Cettolo et al., 2012). Fine-tuning is performed on Multi30k Task 1 (Elliott et al., 2016). That is, importantly, there is no overlap in the pre-training data and the fine-tuning data. Multi30k Task 1 consists of 30k images and one caption per image in English, French, German and Czech (of which we only use the first three). To ensure our findings are robust, we compare four different language models, trained on WikiText103, MS COCO, Flickr30k and all of the above.

The grounding model is trained on Flickr30k (Young et al., 2014). Following Faghri et al. (2018), we randomly crop training images for data augmentation. We use 2048-dimensional features from a pretrained and fixed ResNet-152 (He et al., 2016).

Preprocessing The same tokenization and vocabulary are used across different tasks and datasets. We lowercase and tokenize our corpora with Moses (Koehn et al., 2007) and use subword tokenization with Byte Pair Encoding (BPE) (Sennrich et al., 2016) with 10k merge operations. This allows us to use the same vocabulary across different models seamlessly (translation, language model, image-caption ranker model).

Controling the English message length When fine-tuning the agents, we observe that the length of English messages becomes excessively long. As Agent A has no explicit incentive to output the end-of-sentence (EOS) symbol, it tends to keep transmitting the same token repeatedly. While redundancy might be beneficial for communication, excessively long messages obscure evaluation of the communication protocol. For instance, BLEU score quickly deteriorates as the message length becomes longer, as it is a precision metric. When the message length is fixed, a drop in BLEU score will by necessity mean that the intermediate language has drifted away more. For this reason, we constrain the length of English messages to be no longer than the length of their French source sentence, or shorter if the model outputs the EOS symbol early. Recall that Agent B is supervised to predict the EOS symbol at the right position, so does not suffer from this issue.

Model Architecture and Pretraining Our MT agents are standard sequence-to-sequence models with attention (Bahdanau et al., 2015) with a unidirectional, 1-layer GRU (Cho et al., 2014) with 256 hidden units and 256-dimensional embeddings. During initial pre-training on IWSLT, we early-stop on the validation BLEU score (tst2013). The best checkpoints give 34.05 BLEU and 21.94 BLEU on IWSLT Fr→En and En→De development sets with greedy decoding. For our policy gradient value function, we use a 2-layer MLP with ReLU activations.

The language model is a 1-layer recurrent language model with 512 LSTM hidden units. The image-caption retrieval model is a recently proposed VSE++ model (Faghri et al., 2018), with a unidirectional 1-layer GRU with 512 hidden units and a single fully connected layer from 2048-dimensional ResNet features to 512-dimensional GRU hidden states.

Training Details When fine-tuning our agents, we perform learning rate annealing and early stopping based on Fr→En→De BLEU (communication performance) on the Multi30k development set. We use Adam (Kingma and Ba, 2014) with an initial learning rate of 0.001 and dropout (Srivastava et al., 2014) rate of 0.1. We grid search over the learning rate schedule and the reward coefficients ($\alpha_{\text{pg}}, \alpha_{\text{entr}}, \alpha_{\text{b}}$) for agent A and ($\beta_{LM}, \beta_G$) for agent B, respectively (see previous section). For

our joint systems with policy gradient fine-tuning, we run every model three times with different random seeds and report averaged results.

Baseline and Upper Bound Our main quantitative experiment has three baselines:

- Pretrained models : models pretrained on IWSLT are used without finetuning.

- Ensembling : Given Fr, we let Agent A generate K En hypotheses with beam search. Then, we let Agent B generate the translation $\overline{\text{De}}$ using an ensemble of K source sentences (Firat et al., 2016; Zoph and Knight, 2016).

- Agent A fixed : We fix Agent A (Fr→En) and only fine-tune Agent B using $\mathbb{L}_B$. This shows the communication performance achievable when Agent A cannot drift.

Meanwhile, we also train an NMT model of the same architecture and size directly on the Fr→De task in Multi30k Task 1 (without English intermediary). This serves as an upper bound on the Fr→De performance achievable with available data.

5   Quantitative Results

In Table 2, the top three rows are the baselines described above. The pretrained-only baseline performs relatively poorly on Fr→De, conceivably because it was pretrained on a different corpus in a different domain (IWLST dataset is compiled from TED talks, while Multi30k dataset is a collection of image captions). Ensembling multiple English hypotheses for Agent B gives a negligible increase in Fr→De performance. When only Agent B is fine-tuned and Agent A is kept fixed, we observe an increase from 16.30 to 22.37 in Fr→De. Unsurprisingly, the upper bound NMT model directly trained end-to-end on Multi30k Fr→De (without any pivot, at the bottom of the table) performs best.

When the joint system is fine-tuned on German log-likelihood with policy gradients (PG), we observe a large, 8 BLEU increase in Fr→De (16.30→24.51) at the cost of a substantial, 15 BLEU drop in Fr→En (27.18→12.38). This clearly shows that optimizing for external reward may improve performance on that metric, but at the expense of a drastic language drift in the communication channel on which the reward is imposed.

When the system is fine-tuned only on staying grounded but without any language model con-straint (PG+G), we obtain small performance improvements. This makes sense, since BLEU first and foremost focuses on the surface form. When the agent is trained with the language model constraint (PG+LM), we notice a significant improvement in Fr→En BLEU. When the LM is trained on WikiText103, a widely used language modeling dataset, we observe an improvement of 9 BLEU score over PG (12.38→21.63). When the training corpus is closer to the target domain, such as MS COCO or Flickr30k, we observe even bigger increases. Fr→De translation also improves by 2–3 BLEU (24.51→26.88-27.67).

We see the biggest improvements in performance when agents are trained using both visual grounding feedback and the language model constraint (PG+LM+G). This is particularly pronounced with the LM trained on WikiText103: introducing visual grounding leads to more than 2 BLEU score improvement in Fr→En (21.63→23.65), and 1 BLEU score improvement in Fr→De (26.88→27.87). We hypothesize that the "Englishness" constraint forces agents to communicate with correct syntax and fluency, while the grounding model restricts the search space of languages to ones that are grounded in visual semantics. To investigate the contribution of grounding, we train a much stronger LM on all three datasets combined, and find that there is still more drift even with access to much more language modeling data (23.60→24.75).

It is important to check that the improvement from grounding is actually significant, so we perform a bootstrapped Wilcoxon signed-rank test (Wilcoxon, 1945) on paired English hypotheses for each reference sentence between PG+LM and PG+LM+G, using the model instance that gives the median communication performance (Fr→En→De BLEU) out of three runs. We assess significance on a bootstrapped test set (repeatedly sampled with replacement) and average the statistic over bootstrap samples. With the threshold of $p < 0.02$, PG+LM+G is found to differ significantly for all the LM models, including the All model that had access to much more data. See Table 3.

Figure 2 shows the learning curves, as measured by Fr→De BLEU (left), Fr→En BLEU (middle) and English LM negative log-likelihood (NLL; right). All models improve in fine-tuned task performance (left plot). We observe that vanilla PG fine-tuning quickly leads to highly "un-English" communication, as can be seen from a distinct in-

	LM	Ranker	A: Fr→En	A&B: Fr→En→De
Pretrained			27.18	16.30
Ensembling				16.95
Agent A fixed			27.18	22.37
PG	No LM		12.38 (0.67)	24.51 (1.48)
PG+G	No LM	✔	14.20 (1.58)	26.23 (1.08)
PG+LM	WikiText103		21.63 (1.25)	26.88 (0.12)
	MS COCO		25.05 (1.40)	27.66 (0.34)
	Flickr30k		24.85 (1.14)	27.60 (0.27)
	All		23.60 (1.05)	27.67 (0.39)
PG+LM+G	WikiText103	✔	23.65 (1.91)	27.87 (0.15)
	MS COCO	✔	26.24 (0.28)	27.86 (0.24)
	Flickr30k	✔	25.99 (1.62)	27.82 (0.41)
	All	✔	24.75 (0.40)	28.08 (0.73)
Fr→De				30.73

Table 2: Results in BLEU score on Multi30k Task 1. For our models using policy gradient fine-tuning, we report results averaged over three runs and provide standard deviations in brackets. **PG** (no constraint): trained with vanilla policy gradient fine-tuning. **PG+G** (semantic): trained with grounding only. **PG+LM** (syntactic): trained with "Englishness" constraint. For MS COCO and Flickr30k, the LM was trained directly on image captions. **PG+LM+G** (syntactic+semantic): trained with grounding loss as well as the LM loss. Fr→En: degree of intermediate language drift in agent A; lower indicates more drift. Fr→En→De: overall A&B communication performance; higher is better. For LM=All the LM was trained on all three LM datasets combined.

			PG+LM		
	No LM	WikiText103	MS COCO	Flickr30k	All
+G	✔	✔	✔	✔	✔

Table 3: Using the bootstrapped Wilcoxon signed-rank test (Wilcoxon, 1945), Fr→En results of PG+LM+G are found to be significantly different from its baselines in all cases considered (on all LM datasets) within the threshold of $p = 0.02$.
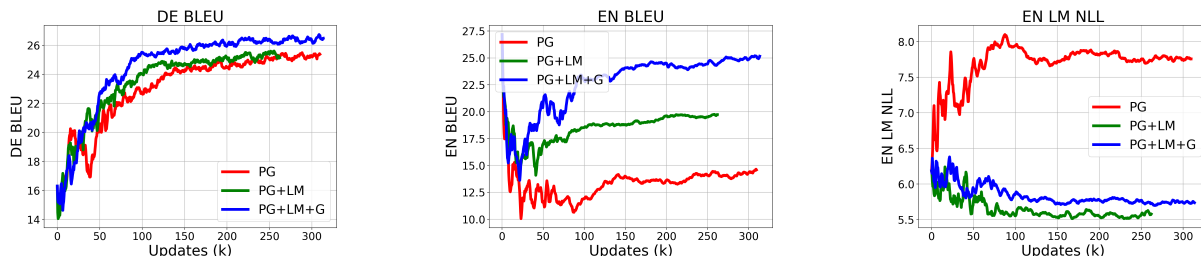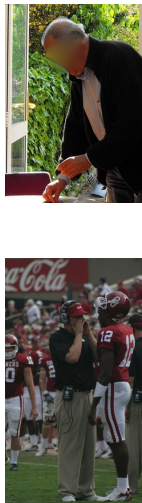

Figure 2: Test set performance over time. En LM NLL curves show the NLL of English messages, computed by a language model trained on WikiText103. Lower En BLEU and higher En LM NLL indicate more language drift.

crease in LM negative log-likelihood (right plot). While PG+LM achieves slightly lower LM NLL than PG+LM+G, its communication protocol drifts much more from English (middle plot). That is, for PG+LM, syntactic conformity is obtained at the expense of semantic preservation. Imposing both syntactic and semantic constraints makes models the least susceptible to drift, almost recovering to the original BLEU score (blue line, middle plot).

6 Analysis

A close investigation into the token statistics of each communication strategy reveals that PG fine-tuning causes the word frequency distribution to be flatter (see Figure 3). The PG model has negative frequency difference values for the most frequent tokens, indicating that PG downweighs frequent words severely, possibly because they are less discriminative. On the other hand, PG+LM gives highly positive frequency differences, meaning that language modeling alone disproportionately emphasizes frequent tokens. Using both the LM and grounding constraints keep the token frequencies closest to the pretrained regimes. Investigating the top 20 most frequent words shows that PG+LM disproportionately favors quotation marks, which are very common in many language modeling datasets but rare in Multi30k (see Table 5).

Ref	Fr	un vieil homme vêtu d'une veste noire regarde sur la table	
	De	ein alter mann in einer schwarzen jacke blickt auf den tisch	
	En	an old man wearing a black jacket is looking on the table	
En	PG	a old teaching black watching on the table table table table table table	
	+LM	a old man in a jacket looking on the table . " "	
	+G	an old man in a black jacket looking on the table .	
De	PG	ein älterer mann in einem schwarzen hemd schaut auf den tisch .	
	+LM	ein alter mann in einer jacke beobachtet einen tisch .	
	+G	ein älterer mann in einer schwarzen jacke schaut auf den tisch .	
Ref	Fr	un joueur de football américain en blanc et rouge parle à un entraîneur .	
	De	ein rot-weiß gekleideter footballspieler spricht mit einem trainer .	
	En	a football player in red and white is talking to a coach .	
En	PG	a player football american football american and red talking talking a coach	
	+LM	a player of white and red talking to a coach . " " "	
	+G	a football player in white and red talking to a coach .	
De	PG	ein footballspieler spricht mit einem spieler in einem roten trikot .	
	+LM	ein weiß gekleideter fußballspieler spricht zu einem trainer .	
	+G	ein fußballspieler in einem rot-weißen trikot spricht mit einem trainer .	

Table 4: Two communication examples on the data from the Multi30k development set with different models (PG, PG+LM, PG+LM+G). The top three rows list the ground truth sentences, the middle three rows are the English messages sent by the Fr→En agent, and the bottom three rows show the German output from the En→De agent. We also show the corresponding images, which are only used to train the grounding model.

Reference	, . the and to of a that i in is it you we 's this "
Pretrained	, the . to of and a i that in it we you 's is this " was
PG	a the and , . in i " this of to is we you ? that not for
PG+LM	the " , of . and in a to this is i es you for we that with
PG+LM+G	the , . of a and to in is i this es we for that you at what

Table 5: Top 20 most frequent tokens (sorted) in English reference (Reference) or the output from Fr→En models.

	IWSLT			Multi30k		
	unique	/sent	/all	unique	/sent	/all
Reference	5,303	19.7	0.86	3,046	11.9	0.91
Pretrained	4,657	17.9	0.85	2,867	12.0	0.87
PG	4,933	13.6	0.56	3,197	9.2	0.65
PG+LM	3,819	14.6	0.61	2,438	10.9	0.78
PG+LM+G	4,327	15.7	0.74	2,550	10.7	0.84

Table 7: Additional token frequency analysis. unique: the number of unique English tokens used in the whole development set. /sent: the number of unique English tokens used per sentence. /all: (the number of unique English tokens / the number of all English tokens.)

	Function words			Content words			
	TO	.	DT	N	V	Adj	Adv
PG	.22	.36	.57	.38	.17	.32	.26
PG+LM	.55	.84	.72	.39	.18	.21	.25
PG+LM+G	.62	.88	.74	.43	.26	.33	.29

Table 6: Exact-match word recall by POS-tag on IWSLT development set: when the English reference contains a word of a certain POS tag, how often does the agent produce it.

Table 6 compares the degree of drift by part-of-speech, and shows that the PG model has very low recall on function words, such as periods and infinitives. Models trained with LM and grounding losses retain function words with much higher accuracy. PG fares relatively better with content words (nouns and verbs), but adding LM and grounding losses still outperform PG. Grounding leads to overall improvements in recall, particularly with content words. Conceivably, when optimizing Agent A's policy on the communication task alone, it is most crucial to relay content information to Agent B, and this might cause agents to ignore syntactic conformity in the original intermediate language. Imposing both syntactic and semantic constraints reduces the space of the intermediate communication protocol to a more stable language space, as reflected in overall task performance.

Table 7 corroborates the finding that vanilla PG fine-tuning leads to flatter token frequency distributions, as the number of unique tokens used by PG is greater than that of the pretrained model. Despite using a more diverse set of tokens, PG uses the smallest number of unique symbols per sentence (/sent) and overall (/all). This implies that PG communication is redundant. PG+LM uses fewer tokens overall, and learns a sharper distribution using a smaller set of high-frequency tokens. Using

Fr src	un enfant assis sur un rocher.	
En ref	a child sitting on a rock formation.	
En hyp	a punk sitting sitting on on a broken	
De ref	ein kind sitzt auf einem felsen .	
De hyp	ein kind sitzt auf einem felsen .	
Fr src	un petit enfant est assis à une table, en train de manger un goûter.	
En ref	a toddler is sitting at a table eating a snack .	
En hyp	a punk sits sitting sitting next next a airline	
De ref	ein kleines kind sitzt an einem tisch und isst einen snack .	
De hyp	ein kind sitzt an einem tisch und liest ein buch .	

Table 8: Evidence of token flipping. The agents use the word "punk" to denote "child" or "baby", which is clearly not desirable.
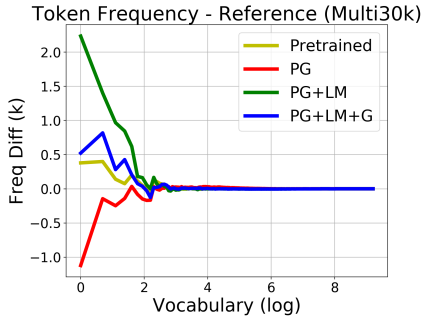

Figure 3: Token frequency analysis on three different models (PG, PG+LM, PG+LM+G) together with the pre-trained model before fine-tuning (Pretrained). We show word frequency curves for each model, after subtracting the reference English frequency statistics (both sorted in decreasing order). Positive y values indicate higher frequency values than the English reference, and negative y values indicate lower frequency values than English. The y-axis is the frequency difference in the thousands, and the x-axis shows the vocabulary index (sorted by frequency) in log scale.

both constraints yields a frequency distribution that most closely resembles the original one.

7 Qualitative Results

In the first example of Table 4 (previous page), it is clear that PG's communication messages have significantly diverged from English: the model is highly repetitive ("table table table table table") and misses some key content words such as "man". Agent B, however, correctly generates the German word "mann". This exemplifies a communication protocol that is successful in solving the task it was trained on, but not fully interpretable to humans. While the output from PG+LM is better, the grounded model's message (PG+LM+G) is distinctly the most fluent and semantically correct.

In the second example, observe that the PG Agent B misinterprets "talking talking a coach a coach" as "spricht mit einem spieler" (talking to a player). The PG+LM+G model again generates a flawless English sentence. Furthermore, its agents succeed in communicating both colors (red and

white) to German while retaining the original English words, when the other models fail to do so.

Interestingly, we observe some instances of token flipping with the PG model and to a lesser extent with the PG+LM model. For example, one particular model uses "punk" to describe "child" (see Table 8). As no occurrence of "punk" in any training data is associated with "child", the agents must have acquired this new meaning assignment during fine-tuning. Among 35 examples in the Multi30k development set where the English reference contains "child", the model uses "punk" 15 times, indicating this is no random phenomenon. We did not observe such examples with the PG+LM+G model.

8 Conclusion

In this paper, we show that language drift happens when fine-tuning natural language agents with some external reward using policy gradients without constraints. We investigate what constraints to put on the communication channel in order to mitigate this. We find that imposing syntactic constraints (via adding language model log-likelihood to the reward) does somewhat mitigate drift, but does not preserve semantic correspondence. We then observe that additionally imposing semantic constraints, e.g. with a perceptual grounding loss, yields communication protocols that best retain the original syntax and intended semantics, while giving the overall best communication performance.

Further analysis into the learned communication protocols reveals that pure PG fine-tuning tends to learn flatter and repetitive token distributions, while encouraging naturalness under a language model disproportionately emphasizes frequent syntactic tokens, yielding a much sharper token distribution than a natural language. The grounded model best retains the original token frequencies.

We examined language drift within a translation game as this allows for direct measurements at each

step (input, intermediate, output), in a way where the semantics stays identical (i.e., the meaning is exactly the same for all languages and modalities) while the communication channel gets only an extrinsic reward (i.e., communication success). The findings in this work, however, are generally applicable to policy gradient fine-tuning of generative language models. We believe that our work shows an intuitive method for addressing language drift and hope that it opens up interesting directions for future work.

Acknowledgments

References

Jacob Andreas, Anca D. Dragan, and Dan Klein. 2017. Translating neuralese. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 232–242.

Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron C. Courville, and Yoshua Bengio. 2016. An actor-critic algorithm for sequence prediction. *arXiv preprint arXiv:1607.07086*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Marco Baroni. 2016. Grounding distributional semantics in the visual world. *Language and Linguistics Compass*, 10(1):3–13.

Ted Briscoe. 2002. *Linguistic evolution through language acquisition*. Cambridge University Press.

Mauro Cettolo, Christian Girardi, and Marcello Federico. 2012. Wit^3: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pages 261–268, Trento, Italy.

Yun Chen, Yang Liu, and Victor O. K. Li. 2018. Zero-resource neural machine translation with multi-agent communication game. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*.

Yong Cheng, Qian Yang, Yang Liu, Maosong Sun, and Wei Xu. 2017. Joint training for pivot-based neural machine translation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 3974–3980.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Grzegorz Chrupała, Akos Kádár, and Afra Alishahi. 2015. Learning language through pictures. *arXiv preprint arXiv:1506.03694*.

Desmond Elliott, Stella Frank, Khalil Sima'an, and Lucia Specia. 2016. Multi30k: Multilingual english-german image descriptions. In *Proceedings of the 5th Workshop on Vision and Language*, pages 70–74.

Desmond Elliott and Akos Kádár. 2017. Imagination improves multimodal translation. *arXiv preprint arXiv:1705.04350*.

Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. 2018. Emergent language in a multi-modal, multi-step referential game. In *Proceedings of the International Conference on Learning Representations*.

Fartash Faghri, David J. Fleet, Jamie Kiros, and Sanja Fidler. 2018. VSE++: improving visual-semantic embeddings with hard negatives. In *British Machine Vision Conference*, page 12.

Orhan Firat, Baskaran Sankaran, Yaser Al-Onaizan, Fatos T Yarman Vural, and Kyunghyun Cho. 2016. Zero-resource translation with multi-lingual neural machine translation. *arXiv preprint arXiv:1606.04164*.

Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 2137–2145.

Jon Gauthier and Igor Mordatch. 2016. A paradigm for situated and goal-driven language learning. *arXiv preprint arXiv:1610.03585*.

Spandana Gella, Rico Sennrich, Frank Keller, and Mirella Lapata. 2017. Image pivoting for learning multilingual multimodal representations. *arXiv preprint arXiv:1707.07601*.

Jiatao Gu, Kyunghyun Cho, and Victor O. K. Li. 2017. Trainable greedy decoding for neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1968–1978.

Serhii Havrylov and Ivan Titov. 2017. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 2146–2156.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 770–778.

Julian Hitschler, Shigehiko Schamoni, and Stefan Riezler. 2016. Multimodal pivots for image caption translation. *arXiv preprint arXiv:1601.03916*.

Ákos Kádár, Desmond Elliott, Marc-Alexandre Côté, Grzegorz Chrupała, and Afra Alishahi. 2018. Lessons learned in multilingual grounded language learning. *arXiv preprint arXiv:1809.07615*.

Douwe Kiela. 2017. *Deep Embodiment: Grounding Semantics in Perceptual Modalities*. Ph.D. thesis, University of Cambridge, Computer Laboratory.

Douwe Kiela, Alexis Conneau, Allan Jabri, and Maximilian Nickel. 2017. Learning visually grounded sentence representations. *arXiv preprint arXiv:1707.06320*.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*.

Simon Kirby. 2001. Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

Satwik Kottur, José M. F. Moura, Stefan Lee, and Dhruv Batra. 2017. Natural language does not emerge 'naturally' in multi-agent dialog. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–2967.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2016. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182*.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-agent cooperation and the emergence of (natural) language. In *Proceedings of the International Conference on Learning Representations*.

Jason Lee, Kyunghyun Cho, Jason Weston, and Douwe Kiela. 2018. Emergent translation in multi-agent communication. In *Proceedings of the International Conference on Learning Representations*.

Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning for negotiation dialogues. *arXiv preprint arXiv:1706.05125*.

Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2157–2169.

Igor Mordatch and Pieter Abbeel. 2017. Emergence of grounded compositional language in multi-agent populations. *arXiv preprint arXiv:1705.10369*.

Kazunori Muraki. 1986. VENUS: two-phase machine translation system. *Future Generation Comp. Syst.*, 2(2):117–119.

Hideki Nakayama and Noriki Nishida. 2017. Zero-resource machine translation by multimodal encoder–decoder network with multimedia pivot. *Machine Translation*, 31(1-2):49–64.

Rodrigo Nogueira and Kyunghyun Cho. 2017. Task-oriented query reformulation with reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 574–583.

Martin A Nowak and David C Krakauer. 1999. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033.

Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

Brian Skyrms. 2010. *Signals: Evolution, learning, and information*. Oxford University Press.

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.

Luc Steels. 1997. The synthetic modeling of language origins. *Evolution of communication*, 1(1):1–34.

Yining Wang, Yang Zhao, Jiajun Zhang, Chengqing Zong, and Zhengshan Xue. 2017. Towards neural machine translation with partially aligned corpora. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pages 384–393.

Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

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.

Barret Zoph and Kevin Knight. 2016. Multi-source neural translation. *arXiv preprint arXiv:1601.00710*.