Emergent Communication: Generalization and Overfitting in Lewis Games

Mathieu Rita
INRIA, Paris
mathieu.rita@inria.fr

Corentin Tallec **Paul Michel*** **Jean-Bastien Grill**
DeepMind
[corentint,paulmiche,jbgrill]@deepmind.com

Olivier Pietquin
Google Research, Brain Team
pietquin@google.com

Emmanuel Dupoux
EHESS,ENS-PSL,CNRS,INRIA
Meta AI Research
emmanuel.dupoux@gmail.com

Florian Strub
DeepMind
fstrub@deepmind.com

Abstract

Lewis signaling games are a class of simple communication games for simulating the emergence of language. In these games, two agents must agree on a communication protocol in order to solve a cooperative task. Previous work has shown that agents trained to play this game with reinforcement learning tend to develop languages that display undesirable properties from a linguistic point of view (lack of generalization, lack of compositionality, etc). In this paper, we aim to provide better understanding of this phenomenon by analytically studying the learning problem in Lewis games. As a core contribution, we demonstrate that the standard objective in Lewis games can be decomposed in two components: a co-adaptation loss and an information loss. This decomposition enables us to surface two potential sources of overfitting, which we show may undermine the emergence of a structured communication protocol. In particular, when we control for overfitting on the co-adaptation loss, we recover desired properties in the emergent languages: they are more compositional and generalize better.

1 Introduction

Understanding the dynamics of language evolution has been a challenging if not controversial research topic in the language sciences [32, 13]. Given that the very first human language cannot be unearthed from fossils [5], computational models have been designed to simulate the emergence of a structured language within a controlled environment. In this line of work, Lewis signaling games [55] are among the most widespread playground environments to model language emergence: they are inherently simple, yet they exhibit a rich set of communication behaviors [17, 70]. Therefore, understanding Lewis games dynamics may shed light on the prerequisites of language emergence.

In their original form, Lewis signaling games involve two agents: a speaker and a listener. The speaker observes a random state from its environment, e.g. an image, and sends a signal to the listener. The listener then undertakes an action based on this signal. Finally, both agents are equally rewarded based on the outcome of the listener's action. The resolution of this cooperative two-player game requires the emergence of a shared protocol between the agents [55, 17]. One way to model the emergence of such protocol is to give the agents the capacity to learn. The agents, and therefore, the communication protocol, are shaped by a sequence of trials and errors over multiple games [81, 44, 75, 70]. This learning-centric approach allows for a fine analysis of the language emergence dynamics [70, 36]. It

*This work was performed when Paul Michel was affiliated with Ecole Normale Supérieure PSL.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

also raises challenging learning-specific questions: What are the inductive biases present in the agent architecture and loss function that shape the emergent language [43]? How do agents generalize from their training set? Is the resulting language compositional [8]? What is the impact of overfitting [53]?

Recently, there has been a resurgence of interest for such learning-based approaches following advances in machine learning [51]. In these approaches, the speakers and listeners are modeled as deep reinforcement learning agents optimized to solve instances of the Lewis games [53, 33, 65, 56, 28]. The vast majority of these works explore Lewis games from an empirical perspective. However, some of the recent experimental results are at odds with experimental findings from the linguistics literature. For instance, the emergent protocols lack interpretability [48], generalization does not always correlate with language compositionality [10], successful strategies are not naturally adopted in populations [68, 12], and anti-efficient communication may even emerge [9]. It is unclear whether those empirical observations result from a learning failure, e.g. optimization problems, overfitting, or whether they are symptomatic of more fundamental limitations of Lewis games for modeling language emergence, e.g. lack of embodiment [31, 4, 63, 37]. Overall, it is crucial to establish new analytical insight to analyze Lewis games in the learning setting.

In this paper, we introduce such an analytical framework to diagnose the learning dynamics of deep reinforcement learning agents in Lewis signaling games. As a core contribution, we demonstrate under mild assumptions that the loss of the speaker and listener can be decomposed into two components when resolving Lewis signaling games: (i) an *information loss* that maximizes the mutual information between the observed states and speaker messages; (ii) a *co-adaptation loss* that aligns the speaker and listener's interpretation of the messages (Section 2). Based on this decomposition, we empirically examine the evolution of these two losses during the learning process (Section 5). In particular, we identify an overfitting problem in the co-adaptation loss between the agents which undermines the emergence of structured language. We then show that the standard setup used in the deep language emergence literature consistently suffers from this overfitting issue (Section 5.1). This realization explains some of the contradictory observations [10] and experimental choices from past works [65, 56, 68]. Finally, we explore regularization methods to tackle this co-adaptation overfitting. We observe that reducing the co-adaptation overfitting allows for developing a more structured communication protocol (Section 5.2).

All in all, our contributions are three-fold: (i) we provide a formal description of Lewis games from a learning standpoint (Section 2.3); (ii) we apply this framework in experiments to show that degenerate results are primarily due to overfitting in the co-adaptation component of the game (Section 5.1) ; (iii) we propose natural ways of tackling this overfitting issue and show that, when we control the receiver's level of convergence, we obtain a well-structured emergent protocol (Section 5.2).

2 Analyzing Lewis Games

We show that Lewis games' objective decomposes into two terms: (i) an information loss that measures whether each message refers to a unique input; (ii) a co-adaptation loss that quantifies the alignment of the speaker's and listener's interpretation of the messages.

For simplicity and to ease the reader's intuition, we focus on the reconstruction variant of Lewis games with agents optimizing the reconstruction log-likelihood in the main paper. In Appendix A, we show that our analysis extends to a broader of Lewis signaling games, e.g. discrimination games [12, 60, 21, 30, 65, 53, 52, 33, 56, 58], and to a general form of reward that covers the rewards commonly used in emergent communication, e.g. log-likelihood [9, 10, 40, 68, 67, 11], accuracy reward [65, 53, 52, 48, 56, 28, 23].

2.1 Background: Lewis Reconstruction Games

Game formalism In reconstruction Lewis games, a speaker observes a random object of its environment. The speaker then sends a descriptive message, which a second agent, the listener, uses to reconstruct the object. The success of the game is quantified by how well the original object is reconstructed [40, 9, 67, 68]. Formally, the speaker is parameterized by θ and the listener is parameterized by ϕ. The observed object denoted by x is selected from a set of objects denoted by $\mathcal{X}$. We denote by X the random variable characterizing x, sampled from distribution p. The intermediate message sent by the speaker m belongs to the set of all potential messages $\mathcal{M}$. The speaker follows

a policy π_θ which samples a message m with probability $\pi_\theta(m|x)$ conditioned on object x. We denote by M_θ the random variable characterizing the message m, sampled from $\pi_\theta(\cdot|X)$. We denote by $\pi_\theta(m) = \sum_x \pi_\theta(m|x)p(x)$ the marginal probability of a message given policy π_θ. Given a message m, the listener outputs a probability distribution over inputs $\rho_\phi(\cdot|m)$, and the probability of reconstructing the entire object x given m is thus $\rho_\phi(x|m)$.

Game objectives In reconstruction games, the speaker and listener minimize the negative log likelihood of the reconstructed object. Both agents thus optimize the objective:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)], \tag{1}$$

where optimizing the speaker is a reinforcement learning problem whose parameters θ are optimized using policy gradient [76] and optimizing the listener is a supervised learning problem whose parameters ϕ are optimized with gradient descent. In our theoretical analysis, we consider that agents are not regularized. In practice, regularizations, e.g. entropy regularization [59], may be added to the game objective but it does not alter our main conclusions.

2.2 Building Intuition on the Lewis Reconstruction Game Learning Dynamics

To get a better intuition of the dynamic of Lewis reconstruction games, we can analyze the form taken by the optimal listener, given speaker π_θ. In what follows, we use sub-script θ to denote an explicit dependency of the policy on parameters θ, e.g. a policy parameterized with a neural network. Conversely, the use of super-script $*(\theta)$ corresponds to an implicit dependency of the policy on parameters θ. As shown in Appendix A.1.1, given a message m, the optimal listener's distribution $\rho^{*(\theta)}(\cdot|m)$ can be written in closed-form:

$$\rho^{*(\theta)}(x|m) := \frac{p(x)\pi_\theta(m|x)}{\sum_{x'\in\mathcal{X}} p(x')\pi_\theta(m|x')}. \tag{2}$$

Here, $\rho^{*(\theta)}$ does not depend on ϕ, but implicitly depends on θ, as it is the optimal listener *given a policy parameterized by* θ. At each update, the listener ρ_ϕ gets closer to its optimum $\rho^{*(\theta)}(\cdot|m)$. If we suppose that the listener perfectly fits $\rho^{*(\theta)}(\cdot|m)$ at any moment, the loss becomes:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[\log \rho^{*(\theta)}(x|m)] = \mathcal{H}(X|M_\theta) = -I(X;M_\theta) + \mathcal{H}(X) \tag{3}$$

where $\mathcal{H}(X|M_\theta)$ is the conditional entropy of X conditioned on M_θ and $I(X;M_\theta)$ is the mutual information between X and M_θ. Thus, if the listener is optimal at every point in time, the speaker's task merely becomes the construction of a message protocol that maximizes the mutual information between objects and messages, i.e. the construction of an unambiguous message protocol.

In practice, the listener never perfectly fits the optimum. In the following, we elucidate the effect of this gap between the listener and its optimum on the dynamics of the game.

2.3 Analytical Result: The Lewis Games Loss Decomposition

In cooperative Lewis games, the agents' loss can be decomposed into two terms:

$$\mathcal{L}_{\theta,\phi} = \mathcal{L}_{\text{info}} + \mathcal{L}_{\text{adapt}}, \tag{4}$$

- An **information term** $\mathcal{L}_{\text{info}}$ quantifies the degree of ambiguity of the language protocol. It is minimal when each message refers to a unique object;
- A **co-adaptation term** $\mathcal{L}_{\text{adapt}}$ quantifies the gap between the listener and its optimum: the speaker's posterior distribution. This co-adaptive term is optimized both by the speaker and the listener. When the listener is optimal, this co-adaptation objective is zeroed.

In particular, the decomposition takes the following form in the Lewis reconstruction game:

$$\mathcal{L}_{\theta,\phi} = \underbrace{\mathcal{H}(X|M_\theta)}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi_\theta} D_{KL}(\rho^{*(\theta)}(\cdot|m)||\rho_\phi(\cdot|m))}_{\mathcal{L}_{\text{adapt}}}, \tag{5}$$

The proof of the decomposition is provided in Appendix A. Appendix A provides the proof for the reconstruction log-likelihood reward and extends to a broader class of Lewis signaling games, e.g. discrimination games, and general cooperative rewards covering usual emergent communication rewards, e.g. the accuracy reward. This decomposition gives us insights on the game dynamics and the constraints that shape languages in the game with neural agents:

The information loss $\mathcal{L}_{\text{info}}$ captures the speaker's intrinsic objective: to develop an unambiguous protocol. $\mathcal{L}_{\text{info}}$ is minimal, equals to 0, when the communication protocol is unambiguous, i.e. every message from the speaker's policy π_θ refers to a unique object. Conversely, $\mathcal{L}_{\text{info}}$ is maximal, equal to $\mathcal{H}(X)$, when the message protocol is fully ambiguous, and X and M_θ are independent variables.

The co-adaptation loss $\mathcal{L}_{\text{adapt}}$ is specific to learning agents. This loss measures how far the listener ρ_ϕ is from its optimum $\rho^{*(\theta)}$. If $\mathcal{L}_{\text{adapt}} = 0$, the listener and its optimum coincide. $\mathcal{L}_{\text{adapt}}$ has the particularity to be optimized by the two agents. From the listener's side, it merely corresponds to the optimization of its supervised task. From the speaker's side, it brings out that the speaker must adapt its language to the listener in addition to build an unambiguous message protocol. In other words, the co-adaptation loss pushes the speaker to develop a language that can be easily recognized by listeners. This pressure diminishes as the listener approaches its optimum.

From a practical perspective, Equation (5) yields the following individual gradients:

$$\begin{cases} \nabla_\theta \mathcal{L}_\theta &= -\nabla_\theta I(X, M_\theta) + \nabla_\theta \mathbb{E}_{m\sim\pi_\theta} D_{KL}(\rho^{*(\theta)}(\cdot|m)||\rho_\phi(\cdot|m)) \\ \nabla_\phi \mathcal{L}_\phi &= \qquad\qquad\qquad\qquad \nabla_\phi \mathbb{E}_{m\sim\pi_\theta} D_{KL}(\rho^{*(\theta)}(\cdot|m)||\rho_\phi(\cdot|m)), \end{cases} \tag{6}$$

where the listener only receives gradients from the co-adaptation term, and the speaker receives gradients from both terms.

This loss decomposition also finds echoes in the cognitive science literature in the form of an expressivity vs. learnability trade-off [72]; see Section 6 for a detailed discussion.

2.4 Generalization Gaps in Lewis Reconstruction Games

We explore another facet of the loss decomposition that arises from learning. As agents are trained on partial views of their environment, it opens questions of overfitting and generalization to unseen objects. As is customary in machine learning, we consider agents trained on a fixed, finite sample from the data distribution: the training set. Let us denote by p_{train} the *empirical* object distribution over the training set and X^{train} an object sampled from p_{train}. Similarly let M_θ^{train} denote a message sampled from $\pi_\theta(.|X^{\text{train}})$, $\pi_\theta^{\text{train}}(m) = \sum_x \pi_\theta(m|x)p_{\text{train}}(x)$ the marginal probability of a message on the training set, and $\rho^{*(\theta)}_{\text{train}}(x|m) = \frac{p_{\text{train}}(x)\pi_\theta(m|x)}{\sum_{x\in\mathcal{X}} p_{\text{train}}(x)\pi_\theta(m|x)}$ the speaker's posterior distribution with respect to the prior distribution p_{train}. The training loss can be written as follow:

$$\begin{aligned} \mathcal{L}^{\text{train}}_{\theta,\phi} &= -\mathbb{E}_{x\sim p_{\text{train}}, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)] \\ &= \underbrace{\mathcal{H}(X^{\text{train}}|M^{\text{train}}_\theta)}_{\mathcal{L}^{\text{train}}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi^{\text{train}}_\theta} D_{KL}(\rho^{*(\theta)}_{\text{train}}(\cdot|m)||\rho_\phi(\cdot|m))}_{\mathcal{L}^{\text{train}}_{\text{adapt}}}. \end{aligned}$$

Decomposing the gap between $\mathcal{L}^{\text{train}}_{\theta,\phi}$ and $\mathcal{L}_{\theta,\phi}$ uncovers two sources of overfitting:

$$\mathcal{L}^{\text{train}}_{\theta,\phi} = \mathcal{L}_{\theta,\phi} + \underbrace{\mathcal{L}^{\text{train}}_{\text{info}} - \mathcal{L}_{\text{info}}}_{\text{information overfitting}} + \underbrace{\mathcal{L}^{\text{train}}_{\text{adapt}} - \mathcal{L}_{\text{adapt}}}_{\text{co-adaptation overfitting}}. \tag{7}$$

Intuitively, *information overfitting* occurs when the speaker only develops an unambiguous language on the training set, but ambiguities remain on the total dataset. *Co-adaptation overfitting* occurs when the two agents agree on a common communication protocol on the training data, but not on all data.

3 Method

This section gathers the methodological tools required to empirically study the loss decomposition.

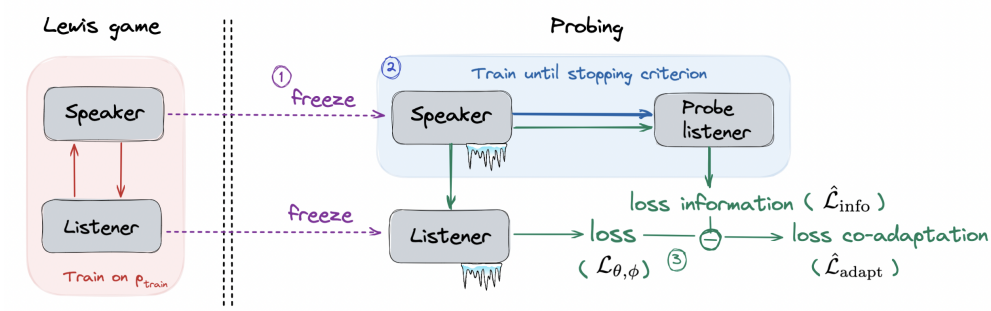

Figure 1: Probing method: (1) the speaker and listener are frozen and the probe listener is initialized. (2) the probe listener is trained on p_{train} (resp. p) with the speaker's messages until convergence; (3) The speaker takes inputs from p_{train} (resp. p_{test}) and messages the probe listener and the listener. The resulting loss of the probe listener is $\hat{\mathcal{L}}_{\text{info}}$, and the loss of the listener is used to estimate $\hat{\mathcal{L}}_{\text{adapt}}$.

3.1 Probing the Information and Co-adaptation Losses

Computing $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$ directly necessitates estimating the posterior distribution of the speaker, $\rho^{*(\theta)}(. \mid m)$. Doing so requires summing over all $\mathcal{X}$ which is intractable. Fortunately, deep models are large enough so that they can perfectly solve their task on their train set. We can leverage this fact to compute empirical estimates $\hat{\mathcal{L}}_{\text{info}}$ and $\hat{\mathcal{L}}_{\text{adapt}}$ of $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$ respectively by using an auxiliary listener trained to optimality.

We here detail an empirical probing mechanism to obtain estimates $\hat{\mathcal{L}}_{\text{info}}$ and $\hat{\mathcal{L}}_{\text{adapt}}$ given speaker π_θ and listener ρ_ϕ. As noted in Equation 2, the posterior $\rho^{*(\theta)}$ also corresponds to the optimal listener. Therefore, we obtain an estimate of the posterior by training a listener to optimality, and use this optimal listener to decompose the loss. In practice, to obtain this optimal listener, we freeze speaker π_θ and listener ρ_ϕ and initialize a new, auxiliary listener from scratch, which we refer to as the *probe* listener. As illustrated in Figure 1, the probe listener is trained to reconstruct object x from message m, with x drawn from distribution p or p_{train} and m sampled according to the frozen speaker policy $\pi_\theta(.|x)$, until a stopping criterion is met. We then distinguish between the train and test estimates:

$$\begin{aligned} \hat{\mathcal{L}}^{\text{train}}_{\text{info}} &= -\mathbb{E}_{x\sim p_{\text{train}}, m\sim\pi_\theta(\cdot|x)}[\log \rho^{\text{train}}_{\omega^*}(x|m)] \\ \hat{\mathcal{L}}^{\text{train}}_{\text{adapt}} &= -\mathbb{E}_{x\sim p_{\text{train}}, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)] - \hat{\mathcal{L}}^{\text{train}}_{\text{info}} \end{aligned} \tag{8}$$

and,

$$\begin{aligned} \hat{\mathcal{L}}^{\text{test}}_{\text{info}} &= -\mathbb{E}_{x\sim p_{\text{test}}, m\sim\pi_\theta(\cdot|x)}[\log \rho_{\omega^*}(x|m)] \\ \hat{\mathcal{L}}^{\text{test}}_{\text{adapt}} &= -\mathbb{E}_{x\sim p_{\text{test}}, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)] - \hat{\mathcal{L}}^{\text{test}}_{\text{info}} \end{aligned} \tag{9}$$

where $\rho^{\text{train}}_{\omega^*}$ and ρ_{ω^*} are the probe listeners trained over distributions p_{train} and p respectively.[2] Note that this probing mechanism, while tractable, is computationally costly as it necessitates training a new probe listener to convergence, and so we only use it as a valuable diagnosis tool.

3.2 Balancing the Information and Co-adaptation Terms

As explained in Section 2.3, the information loss alone is sufficient for the speaker to develop an unambiguous language. This begets the question: does the co-adaptation loss have any bearing on the emergent language at all? We elucidate this question by balancing the weight of the co-adaptation term in the decomposition. By using the probing method described above, we build the following training loss:

$$\mathcal{L}_\theta(\alpha) = (1-\alpha) \times \hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \alpha \times \hat{\mathcal{L}}^{\text{train}}_{\text{adapt}} \qquad \text{where} \quad \alpha \in [0; 0.5] \tag{10}$$

[2]The estimate is trained on p, the *full* distribution of objects, and not p_{test}. Training on p_{test} would result in an optimal listener overfitting on the test set, which would results in bad estimates of the mutual information.

Hence, α balances the two speaker objectives (up to an approximation error). When $\alpha = 0.5$, the loss falls back to the classic setting. When $\alpha = 0$, the co-adaptation term is removed on the speaker side; note that the Lewis game can still be solved since the listener still optimizes the co-adaptation term. We experimentally analyse the effect of α on resulting languages in Section 5.1. In Appendix B, we describe how we build the balanced loss and explain why α should be bounded by 0.5.

3.3 Controlling the Listener's Co-adaptation Loss Level of Convergence

As mentioned in 2.2, the influence of $\mathcal{L}_{\text{adapt}}$ on the co-adaptation term in the speaker's loss is modulated by the listener's level of convergence to its optimum. To understand the effect of this co-adaptation, we decouple the speaker and listener training and train the listener via three procedures:

Continuous listener The listener is continuously trained, jointly with the speaker. This is the standard setting in the emergent communication literature, and serves to report the baseline behavior.

Partial listener The listener is re-initialized *after each* of the speaker's update and trained on the training set for N_{step} before updating the speaker again. This baseline enables fine-grained analysis of the influence of under-training (low N_{step}) and over-training (large N_{step}) the listener.

Early stopping listener The listener is also re-initialized *after each* of the speaker's update but is now trained until an early stopping criterion is met on the validation set. This allows us to get the best estimate of the posterior $\rho^{*(\theta)}(.|m)$ at each update. This can be seen as a variant of the partial listener with an adaptive number of steps N_{step}.

4 Experimental settings

4.1 Game description

Unless specified, all our experiments are run on the reconstruction game defined in Section 2.1. Experiments are run over 6 seeds and reach $> 99\%$ training reconstruction scores unless otherwise stated. Our implementation is based on the EGG toolkit [39] and the code is available at `https://github.com/MathieuRita/Population`.

Environment We consider objects $x =: (x_1, ..., x_K) \in \mathcal{X} =: \mathcal{X}_1 \times ... \times \mathcal{X}_K$ characterized by K attributes where attribute i may take $|\mathcal{X}_i|$ different values. By design, this synthetic environment allows us to test the ability of agents to refer to unseen objects by communicating their attributes [3, 48]. Each object is the concatenation of one-hot representations of the attributes $(x_i)_{1 \leq i \leq K}$. Objects have $K = 6$ attributes, each taking 10 different values, for a total of 1 million objects. Training, validation and test sets are randomly drawn from this pool of objects (uniformly and without overlap), and are respectively composed of 4000, 1000 and 1000 elements. Thus, the agents only have access to a small fraction ($< 1\%$) of the environment, making the generalization problem challenging.

Communication channel Messages $m =: (m_j)_{j=1}^T \in \mathcal{M} =: \mathcal{V}^T$ are sequences of T tokens where each token is taken from a finite vocabulary $\mathcal{V}$, finishing by a hard-coded end-of-sentence token `EoS`. In our experiments, messages have maximum length $T = 10$ and symbols are taken from a vocabulary of size $|\mathcal{V}| = 10$ to prevent a bottleneck in the communication channel.

Speaker model The speaker follows a recurrent policy: given an input object x, it samples for all $t \in [1, T]$ a token m_t with probability $\pi_\theta(m_t|m_{<t}, x)$. The speaker takes in the object x as a vector of size $K \times |\mathcal{X}_.|$ and passes it through a linear layer of size 128 to obtain an object embedding, used to initialize a LSTM [35] of size 128 with layer normalization [2]. At each time step, the LSTM's output is fed into a linear layer of size $|\mathcal{V}|$, followed by a softmax, to produce $\pi_\theta(m_t|m_{<t}, x)$

Listener model Given a message $m = (m_1, ..., m_T)$, the listener outputs for each attribute k a probability distribution over the $|\mathcal{X}_k|$ values: $\rho_\phi^k(\cdot|m)$. The probability of reconstructing the entire object x given m is then $\rho_\phi(x|m) := \prod_k \rho_\phi^k(x^k|m)$. The listener passes each message m_t through an embedding layer of dimension 128 followed by a LSTM with layernorm of size 128. The final recurrent state h_T^1 is passed through K linear projections of size $|\mathcal{X}_.|$, each followed by a softmax, providing K independent probability distributions of sizes $|\mathcal{X}_.|$ to predict each attribute of x.

Optimization The agents are optimized using `Adam` [42] with a learning rate of $5 \cdot 10^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and a batch size of 1024. For the speaker we use policy gradient [76], with a baseline

computed as the average reward within the minibatch, and an entropy regularization of 0.01 to the speaker's loss [82]. In all experiments, we select the best models by early stopping.

4.2 Evaluating emergent languages properties

Generalization We measure generalization by computing the average test reconstruction score over all the attributes of a probe listener trained on the training set using an early stopping criterion on the validation set. Indeed, the trained listener ρ_ϕ may overfit to the training set, and so using it may under-estimate. Using a separate listener removes this bias.

Compositionality Compositionality is a fundamental feature of natural language often seen as a precondition to generalize [6, 77, 79]. We assess the compositionality by computing the topographic similarity [8, 53]. It is defined as the Spearman correlation [47, 80] between the distance in input space, i.e. the average number of common attributes, and the distance in message space, i.e. the edit-distance between the corresponding messages [54]. As we here deal with large object space and stochastic policies, we use a bootstrapped estimate of topographic similarity as in [46] to get reliable numbers. We sub-sample 1000 elements x from the object space $\mathcal{X}$, and sample the corresponding message m from the speaker's policy $\pi_\theta(\cdot|x)$. We compute the topographic similarity for this batch of 1000 pairs (x, m). We repeat this protocol 100 times and take the mean to measure compositionality.

5 Empirical results

5.1 Visualizing the loss decomposition dynamics

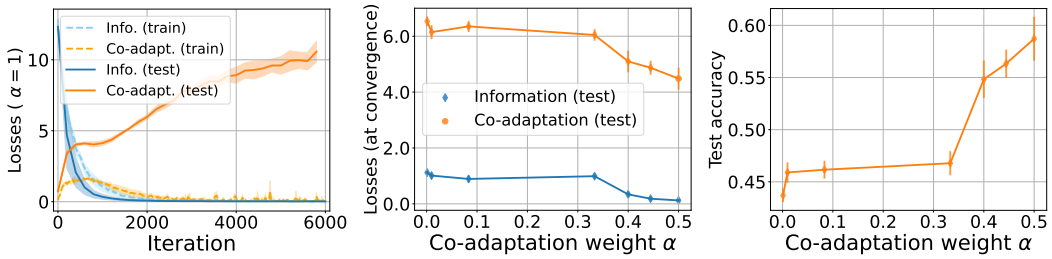

Figure 2: (a)Training dynamics ($\alpha = 0.5$). (b,c)Agents score as a function of co-adaptation weight α.

We here visualize the loss decomposition dynamics. Following the protocol of 3.2, we control $\mathcal{L}^{\text{train}}_{\text{adapt}}$ in speaker's loss with weight α to understand the influence of the co-adaptation term on the language.

The co-adaptation task overfits rapidly We plot information and co-adaptation training dynamics in the standard setting ($\alpha = 0.5$). Note that both train and test information losses quickly converge to 0, in other words the speaker succeeds in developing a protocol that is unambiguous on both the training set and the overall distribution. On the other hand, the test co-adaptation loss diverges while the train co-adaptation keeps disminishing, highlighting a clear overfitting problem.

The co-adaptation task promotes generalization We then display in Figure 2 the evolution of the information and co-adaptation losses for different co-adaptation weight α. We observe that down-weighting $\mathcal{L}^{\text{train}}_{\text{adapt}}$ tends to enforce both information and co-adaptation overfitting. Thus, even though the co-adaptation loss is not inherently necessary for the speaker to develop an unambiguous language, it is important to encourage the speaker to build a better language. This is confirmed when looking at generalization accuracies. From $\alpha = 0$ to $\alpha = 0.5$, there is a gain of 15 points of generalization. In conclusion, we note that (i) balancing the loss in favor of $\mathcal{L}^{\text{train}}_{\text{info}}$ has a negative impact on generalization, (ii) the co-adaptation loss $\mathcal{L}^{\text{train}}_{\text{adapt}}$ pushes the speaker to develop a language that generalizes better.

These experiments highlight two key findings: (i) co-adaptation is crucial for generalization ; (ii) in standard settings, the co-adaptation loss overfits substantially, whereas the information loss does not.

5.2 Countering co-adaptation overfitting

We here investigate whether limiting overfitting in the co-adaptation loss may push towards languages that generalize better and are more stuctured. As described in 3.3, we compare three control baselines: *Continuous listener*, *Partial listener* with varying levels of convergence, and *Early stopping listener*.

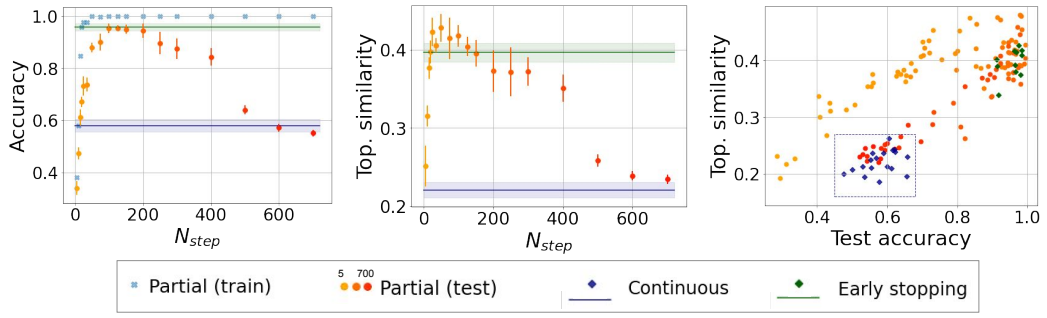

Figure 3: (a,b) Evolution of generalization and top.sim with *Partial listener*'s number of learning steps N_{step} ; (c) Top. sim VS. generalization. The color level of orange dots increases with N_{step}. Blue (resp. green) lines and points refer to the Continuous listener (resp. Early stopping listener).

Countering co-adaptation overfitting improves generalization In Figure 3, we observe that the level of convergence of the *Partial listener* between each speaker's update (controlled by N_{step}) has a strong impact on the generalization of the emergent protocol. Overall, we recover classic machine learning trends when varying N_{step}: when $N_{step} < 50$, both train and test accuracy are low — the agents underfit. When $50 < N_{step} < 250$, the train and test accuracy are almost optimal — the agents are in good training regime. Finally, when $N_{step} > 250$, the train accuracy is maximal while the test accuracy collapses — the agents overfit. These observations reveal that the level of convergence of the listener has a substantial impact on the final emergent language capacity to generalize. Recall that, in these experiments, the direct effect of the listener's overfitting is mitigated, as we measure generalization using an auxiliary listener that is early stopped, and should therefore not overfit as noted in Section 4. The listener's overfitting impacts the speaker's update through the co-adaptation loss, which, by inducing a poorer final language leads to a degradation in generalization. Additionally, Figure 3 shows that the continuous listener, standard in the Lewis games literature, provides generalization performance similar to the worst overfitting listeners.

Controlling the listener's co-adaptation level appears crucial to let the speaker develop a language that generalizes well; this effect may have been underestimated in the standard Lewis learning dynamic.

Countering co-adaptation overfitting improves compositionality Figure 3 reveals that compositionality follows the same pattern. In the underfitting regime, the topographic similarity is low but still outperforms the *Continuous listener*. Similarly, it is also low in the overfitting regime. In-between the two — which corresponds to high generalization in Figure 3 — the topographic similarity reaches high values, which suggests that more compositional languages emerge. This indicates that the listener's lack of co-adaptation overfitting promotes structured languages.

Compositionality correlates with generalization In Figure 3, we plot the correlation between generalization and compositionality. As opposed to [10], we observe a strong correlation between generalization and topographic similarity when varying the *Partial listener*'s level of convergence. In particular, we identify two correlation branches: one belonging to the underfitting regime and the second to the overfitting regime. Together, they retrace the evolution of generalization and compositionality with respect to N_{step}. We see that *Continuous listeners* belong to the end of this trajectory, in the overfitting regime. Note that the blue rectangle — which delineates the range of values reached with the *Continuous listener* — corresponds to the classic learning setting in the literature. As this range is tight, it may explain the initial negative results reported by [10].

In conclusion, the listener exerts a necessary pressure on the speaker to develop a structured language that generalizes better. This pressure can be controlled by limiting the listener's level of overfitting, which is inevitably too high when the listener is trained continuously as is usually done.

Comparison with standard regularization methods In practice, re-initializing the listener as done with the *Partial* or *Early stopping listener* is costly. We thus test whether performances comparable to Figure 3 can be obtained by controlling the listener's level of overfitting with standard regularization methods. In Table 1, we report the influence of applying common regularization methods to the listener on various metrics of the language. We find that regularization consistently results in noticeable improvements. Moreover, once again, gains of generalization correlate with gains of

	Gen. ↑	Compo. ↑	$\hat{\mathcal{L}}^{\text{test}}_{\text{adapt}}$ ↓
Continuous	$0.58_{\pm 0.05}$	$0.22_{\pm 0.02}$	$4.64_{\pm 1.22}$
Dropout	$0.64_{\pm 0.03}$	$0.24_{\pm 0.01}$	$4.86_{\pm 0.52}$
No LN.	$0.70_{\pm 0.03}$	$0.24_{\pm 0.02}$	$4.68_{\pm 0.38}$
Weight decay	$0.72_{\pm 0.03}$	$0.25_{\pm 0.03}$	$4.29_{\pm 0.56}$
No LN. + WD	$0.87_{\pm 0.07}$	$0.30_{\pm 0.03}$	$2.12_{\pm 0.67}$
Early stopping	$\mathbf{0.95}_{\pm \mathbf{0.04}}$	$0.39_{\pm 0.04}$	$1.10_{\pm 0.69}$
Top Partial	$\mathbf{0.95}_{\pm \mathbf{0.03}}$	$\mathbf{0.42}_{\pm \mathbf{0.02}}$	$\mathbf{0.97}_{\pm \mathbf{0.55}}$

	Generalization ↑		
CelebA	1/1	1/20	1/100
Continuous	$0.94_{\pm 0.01}$	$0.67_{\pm 0.02}$	$0.39_{\pm 0.07}$
Early stopping	$\mathbf{0.97}_{\pm \mathbf{0.01}}$	$\mathbf{0.80}_{\pm \mathbf{0.03}}$	$\mathbf{0.69}_{\pm \mathbf{0.04}}$
ImageNet	1/1	1/20	1/100
Continuous	$0.96_{\pm 0.01}$	$0.77_{\pm 0.01}$	$0.51_{\pm 0.03}$
Early stopping	$\mathbf{0.98}_{\pm \mathbf{0.01}}$	$\mathbf{0.81}_{\pm \mathbf{0.01}}$	$\mathbf{0.64}_{\pm \mathbf{0.01}}$

Table 1: (left) Performance comparisons between Continuous listener, Partial listener, Early stopping listener and classic listener regularization, e.g. weight decay [34, 49], Dropout [74] and layernorm [2]. Regularization parameters were tuned and are detailed in Appendix C.1 ; (right) Generalization scores for continuous baselines and Early stopping listener on visual Lewis Games. $1/1$, $1/20$ and $1/100$ refer to the subset ratios of the dataset.

compositionality. These trends corroborate our hypothesis that controlling the listener's learning is key to encourage the speaker to develop more structured languages. However, those methods remain under the upper bound reached by the *Early stopping listener*, which suggests that further research on regularization in cooperative games is warranted.

We complement this analysis in Appendix C.2 by studying the impact of regularization on the speaker's side, and show that such regularization does not result in similar improvements. This indicates that the listener is the main contributor to the co-adaptation overfitting.

5.3 Scaling to the Image Discrimination Games

To validate our empirical findings beyond synthetic games, we scale our approach to complex games with natural images as advocated by [12]. We thus train our agents on a discriminative game on top of the CelebA [57] and ImageNet [69, 19] datasets while applying previous protocol. We work on 3 sizes of training set with increasing generalization difficulty. We provide all the training details and game settings in Appendix D.1 and report our results in Table 1. While agents generalize well when trained on the entire training set, generalization issues occur on smaller training sets and performances can indeed be improved by controlling the listener's level of convergence. However, Appendix D.2 shows that gain of generalization does not correlate with gain of topographic similarity, supporting that agents' language structure is not captured by the topographic similarity in image based settings [12, 1].

6 Related work

The decomposition of the loss function in the Lewis Game that we introduced finds echos in the cognitive science literature. According to Skyrms [70], communicative organisms or systems are confronted with two types of information: about the environmental states shared by the agents (called *objective* information), and about how an agent would react to a signal (called *subjective* information). Communication protocols emerge as a trade-off between constraints related to those two types of information [44, 45]: the sender should be expressive [25, 24] and transcribe the information available in the world with as little ambiguity as possible, which has been described as a *bias against ambiguity* [73] ; sender and receiver should agree on the same referring system, which has been described as a *conceptual pact* [7]. The latter has been shown to impose compressibility and learnability pressures promoting structure [78, 72, 84]. This analysis resonates well with our analytical decomposition of the loss function in the Lewis game.

The first term of the decomposition, which we called the information loss, has been addressed by previous work that assumed that linguistic structure and generalization emerge from the requirement of creating an unambiguous language. In this line of work, studies have either manipulated the complexity of the environment [12, 30, 71, 60], restricted the bandwidth of the communication channel [48, 66], or added noise to the message [50, 85]. In our main experiment, we do not apply such information constraints to better focus on the second term of the decomposition, the *co-adaptation* constraint, less studied within a machine learning approach. Previous work have assumed that the co-adaptive dynamics encourage speakers to develop a more structured language for

learnability reasons [56]. Support for this hypothesis can be found directly via the implementation of a neural variant of Iterated Learning [65] or the introduction of learning speed heterogeneities [68] and indirectly via the restriction of agents capacity [66], the variation of the communication-graph in populations [28, 41] or the addition of newborn agents [15]. In our paper, we demonstrate that a co-adaptation term is always present in standard agents optimization protocols and show that controlling *co-adaptation overfitting* enhances language properties. The existence of an overfitting regime found under the default setting (continuous training) may explain the counter-intuitive lack of relationship between compositionality and generalization previously reported with neural agents [51, 10, 38, 20].

7 Conclusion

In this paper, we propose a methodological approach to better understand the dynamics in Lewis signaling games for language emergence. It allows us to surface two components of the training: (i) an information loss, (ii) a co-adaptation loss. We shed light that the agents tend to overfit this co-adaptation term during training, which hinders the learning dynamic and degrades the resulting language. As soon as this overfitting is controlled, agents develop compositional languages that better generalize. Remarkably, this emergent compositionality does not result from environmental factors, e.g. communication bottleneck [43], under-parametrization [48, 26], population dynamics [12, 68], memory restriction [15, 16] or inductive biases [67], but only through a trial-and-error process. Therefore, we advocate for a better comprehension of the optimization and machine learning issues. As illustrated in this paper, such understanding may unveil contradictions between computational models and language empirical observations and better expose the existing synergies between learning dynamics and environmental factors [27, 83, 64, 14, 18, 22].

Acknowledgments

Authors would like to thank Rahma Chaabouni, Marco Baroni, Paul Smolensky, Bilal Piot and Karl Tuyls for helpful discussions and the anonymous reviewers to their relevant comments. M.R would also like to thank Michael Sander and Maureen de Seyssel for last minute feedbacks. M.R. was supported by the MSR-Inria joint lab and granted access to the HPC resources of IDRIS under the allocation 2021-AD011012278 made by GENCI. P.M. was supported by the ENS-CFM Data Science Chair. E.D. was funded in his EHESS role by the European Research Council (ERC-2011-AdG-295810 BOOTPHON), the Agence Nationale pour la Recherche (ANR-17-EURE-0017 Frontcog, ANR-10-IDEX0001-02 PSL*, ANR-19-P3IA-0001 PRAIRIE 3IA Institute) and grants from CIFAR (Learning in Machines and Brains) and Meta AI Research (Research Grant).

References

[1] Jacob Andreas. Measuring compositionality in representation learning. In *Proc. of International Conference on Learning Representations (ICLR)*, 2019.

[2] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[3] Marco Baroni. Linguistic generalization and compositionality in modern artificial neural networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791):20190307, 2020. doi: 10.1098/rstb.2019.0307.

[4] Lawrence W Barsalou. Grounded cognition. *Annu. Rev. Psychol.*, 59:617–645, 2008.

[5] Derek Bickerton. Language evolution: A brief guide for linguists. *Lingua*, 117(3):510–526, 2007.

[6] Derek Bickerton. More than nature needs. In *More than Nature Needs*. Harvard University Press, 2014.

[7] Susan E Brennan and Herbert H Clark. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, 22(6):1482, 1996.

[8] Henry Brighton and Simon Kirby. Understanding Linguistic Evolution by Visualizing the Emergence of Topographic Mappings. *Artificial Life*, 12(2):229–242, 04 2006. ISSN 1064-5462. doi: 10.1162/artl.2006.12.2.229.

[9] Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Anti-efficient encoding in emergent communication. In *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[10] Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. Compositionality and generalization in emergent languages. In *Proc. of the Association for Computational Linguistics (ACL)*, 2020.

[11] Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12), 2021.

[12] Rahma Chaabouni, Florian Strub, Florent Altché, Eugene Tarassov, Corentin Tallec, Elnaz Davoodi, Kory Wallace Mathewson, Olivier Tieleman, Angeliki Lazaridou, and Bilal Piot. Emergent communication at scale. In *Proc. of International Conference on Learning Representations (ICLR)*, 2022.

[13] Morten H Christiansen and Simon Kirby. Language evolution: Consensus and controversies. *Trends in cognitive sciences*, 7(7):300–307, 2003.

[14] Michael Clyne. Linguistic and sociolinguistic aspects of language contact, maintenance and loss. *Maintenance and loss of minority languages*, 1:17, 1992.

[15] Michael Cogswell, Jiasen Lu, Stefan Lee, Devi Parikh, and Dhruv Batra. Emergence of compositional language with deep generational transmission. *arXiv preprint arXiv:1904.09067*, 2019.

[16] Hannah Cornish, Rick Dale, Simon Kirby, and Morten H Christiansen. Sequence memory constraints give rise to language-like structure through iterated learning. *PloS one*, 12(1):e0168532, 2017.

[17] Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

[18] General Cultural Intelligence Team, Avishkar Bhoopchand, Bethanie Brownfield, Adrian Collister, Agustin Dal Lago, Ashley Edwards, Richard Everett, Alexandre Frechette, Yanko Gitahy Oliveira, Edward Hughes, et al. Learning robust real-time cultural transmission without human data. *arXiv preprint arXiv:2203.00715*, 2022.

[19] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[20] Roberto Dessì and Marco Baroni. Cnns found to jump around more skillfully than rnns: Compositional generalization in seq2seq convolutional networks. *arXiv preprint arXiv:1905.08527*, 2019.

[21] Roberto Dessì, Eugene Kharitonov, and Baroni Marco. Interpretable agent communication from scratch (with a generic visual processor emerging on the side). *Advances in Neural Information Processing Systems*, 34, 2021.

[22] Nick C Ellis. The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The modern language journal*, 92(2):232–249, 2008.

[23] Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. Emergent communication in a multi-modal, multi-step referential game. In *Proc. of International Conference on Learning Representations (ICLR)*, 2018.

[24] Nicolas Fay and T Mark Ellison. The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one*, 8(8):e71781, 2013.

[25] Bruno Galantucci and Simon Garrod. Experimental semiotics: a review. *Frontiers in human neuroscience*, 5:11, 2011.

[26] Lukas Galke, Yoav Ram, and Limor Raviv. Emergent communication for understanding human language evolution: What's missing? In *Emergent Communication Workshop at ICLR*, 2022.

[27] Rick Dale Gary Lupyan. Language structure is partly determined by social structure. *PLoS ONE 5*, 1, 2010.

[28] Laura Graesser, Kyunghyun Cho, and Douwe Kiela. Emergent linguistic phenomena in multi-agent communication games. In *Proc. of Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[29] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in Neural Information Processing Systems*, 33:21271–21284, 2020.

[30] Shangmin Guo, Yi Ren, Kory Mathewson, Simon Kirby, Stefano V Albrecht, and Kenny Smith. Expressivity of emergent language is a trade-off between contextual complexity and unpredictability. In *Proc. of International Conference on Learning Representations (ICLR)*, 2021.

[31] Stevan Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346, 1990.

[32] Steven R Harnad, Horst D Steklis, and Jane Ed Lancaster. Origins and evolution of language and speech. *Annals of the New York Academy of Sciences*, 1976.

[33] Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[34] Geoffrey E Hinton. Learning translation invariant recognition in a massively parallel networks. In *International Conference on Parallel Architectures and Languages Europe*, pages 1–13. Springer, 1987.

[35] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[36] Simon Huttegger, Brian Skyrms, Pierre Tarres, and Elliott Wagner. Some dynamics of signaling games. *Proceedings of the National Academy of Sciences*, 111(Supplement 3):10873–10880, 2014.

[37] Aleksandra Kalinowska, Elnaz Davoodi, Florian Strub, Kory Mathewson, Todd Murphey, and Patrick Pilarski. Situated communication: A solution to over-communication between artificial agents. In *Emergent Communication Workshop at ICLR 2022*, 2022.

[38] Eugene Kharitonov and Marco Baroni. Emergent language generalization and acquisition speed are not tied to compositionality. In *Proc. of the BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 2020.

[39] Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. Egg: a toolkit for research on emergence of language in games. In *Proc. of Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[40] Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. Entropy minimization in emergent languages. In *Proc. of International Conference on Machine Learning (ICML)*, 2020.

[41] Jooyeon Kim and Alice Oh. Emergent communication under varying sizes and connectivities. *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[42] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. of International Conference on Learning Representations (ICLR)*, 2015.

[43] Simon Kirby. 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, 2001.

[44] Simon Kirby and James R Hurford. The emergence of linguistic structure: An overview of the iterated learning model. *Simulating the evolution of language*, pages 121–147, 2002.

[45] Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102, 2015.

[46] Philipp Koehn. Statistical significance tests for machine translation evaluation. In *Proc. of Empirical Methods in natural Language Processing (EMNLP)*, 2004.

[47] Stephen Kokoska and Daniel Zwillinger. *CRC standard probability and statistics tables and formulae*. Crc Press, 2000.

[48] Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra. Natural language does not emerge 'naturally' in multi-agent dialog. In *Proc. of Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

[49] Anders Krogh and John Hertz. A simple weight decay can improve generalization. *Advances in neural information processing systems*, 4, 1991.

[50] Łukasz Kuciński, Tomasz Korbak, Paweł Kołodziej, and Piotr Miłoś. Catalytic role of noise and necessity of inductive biases in the emergence of compositional communication. *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[51] Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*, 2020.

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

[53] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. In *Proc. of International Conference on Learning Representations (ICLR)*, 2018.

[54] Vladimir I Levenshtein et al. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10:8, pages 707–710. Soviet Union, 1966.

[55] David Kellogg Lewis. *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell, 1969.

[56] Fushan Li and Michael Bowling. Ease-of-teaching and language structure from emergent communication. In *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[57] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proc. of the International Conference on Computer Vision (ICCV)*, 2015.

[58] Ryan Lowe, Abhinav Gupta, Jakob Foerster, Douwe Kiela, and Joelle Pineau. On the interaction between supervision and self-play in emergent communication. In *Proc. of International Conference on Learning Representations (ICLR)*, 2020.

[59] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1928–1937, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/mniha16.html.

[60] Jesse Mu and Noah Goodman. Emergent communication of generalizations. *Proc. of Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[61] Michael Noukhovitch, Travis LaCroix, Angeliki Lazaridou, and Aaron Courville. Emergent communication under competition. In *Proc. of International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, 2021.

[62] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[63] Marie Ossenkopf, Kevin Sebastian Luck, and Kory Wallace Mathewson. Which language evolves between heterogeneous agents? - communicating movement instructions with widely different time scopes. In *Emergent Communication Workshop at ICLR 2022*, 2022.

[64] Limor Raviv, Antje Meyer, and Shiri Lev-Ari. Compositional structure can emerge without generational transmission. *Cognition*, 182:151–164, 2019. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2018.09.010. URL https://www.sciencedirect.com/science/article/pii/S0010027718302464.

[65] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. In *Proc. of International Conference on Learning Representations (ICLR)*, 2020.

[66] Cinjon Resnick, Abhinav Gupta, Jakob Foerster, Andrew M. Dai, and Kyunghyun Cho. Capacity, bandwidth, and compositionality in emergent language learning. In *Proc. of Autonomous Agents and Multiagent Systems (AAMAS)*, 2020.

[67] Mathieu Rita, Rahma Chaabouni, and Emmanuel Dupoux. "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In *Proc. of the Conference on Computational Natural Language Learning (CoNLL)*, 2020.

[68] Mathieu Rita, Florian Strub, Jean-Bastien Grill, Olivier Pietquin, and Emmanuel Dupoux. On the role of population heterogeneity in emergent communication. In *Proc. of International Conference on Learning Representations (ICLR)*, 2022.

[69] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision (IJCV)*, 115(3):211–252, 2015.

[70] Brian Skyrms. *Signals: Evolution, learning, and information*. OUP Oxford, 2010.

[71] Agnieszka Słowik, Abhinav Gupta, William L Hamilton, Mateja Jamnik, Sean B Holden, and Christopher Pal. Exploring structural inductive biases in emergent communication. *arXiv preprint arXiv:2002.01335*, 2002.

[72] Kenny Smith, Simon Kirby, and Henry Brighton. Iterated learning: A framework for the emergence of language. *Artificial life*, 9(4):371–386, 2003.

[73] Matthew Spike, Kevin Stadler, Simon Kirby, and Kenny Smith. Minimal requirements for the emergence of learned signaling. *Cognitive science*, 41(3):623–658, 2017.

[74] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56): 1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

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

[76] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*, 2000.

[77] Zoltán Gendler Szabó. Compositionality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition, 2020.

[78] Mónica Tamariz and Simon Kirby. Culture: copying, compression, and conventionality. *Cognitive science*, 39(1):171–183, 2015.

[79] Simon W Townsend, Sabrina Engesser, Sabine Stoll, Klaus Zuberbühler, and Balthasar Bickel. Compositionality in animals and humans. *PLoS Biology*, 16(8):e2006425, 2018.

[80] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[81] Kyle Wagner, James A Reggia, Juan Uriagereka, and Gerald S Wilkinson. Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37–69, 2003.

[82] Ronald J Williams and Jing Peng. Function optimization using connectionist reinforcement learning algorithms. *Connection Science*, 3(3):241–268, 1991.

[83] Alison Wray and George W. Grace. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3):543–578, 2007. ISSN 0024-3841. The Evolution of Language.

[84] Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proc. of the National Academy of Sciences*, 115(31):7937–7942, 2018. doi: 10.1073/pnas.1800521115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1800521115.

[85] Łukasz Kuciński, P. Kołodziej, and Piotr Milos. Emergence of compositional language in communication through noisy channel. In *Language in Reinforcement Learning (LaReL), ICML Workshop*, 2020.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We mentioned the Partial and Early Stopping listener training are costly opening a discussion on regularization methods (Section 5.2)
   (c) Did you discuss any potential negative societal impacts of your work? [No] We run experiments that have no negative societal impacts
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] We took a large part of our analytical part to detail the set of assumptions. Further discussions to more general cases are discussed in Appendix A
   (b) Did you include complete proofs of all theoretical results? [Yes] If not in the main paper, all derivations are brought in the Appendices
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] Code will be released upon deanonymization.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] Our work was carried out on GPUs on an institutional cluster. It is will be mentionned upon deanonymization. Each experiment runs on a single V100-32G GPU
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] The ImageNet and CelebA papers were cited
   (b) Did you mention the license of the assets? [No]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Lewis Games Loss Decomposition : proofs

We provide all the proofs of the Lewis Games Loss Decomposition. We organize the proofs as follows:

- **Appendix A.1 - Reconstruction game**, we provide the proofs of the Decomposition for the reconstruction game.
  - **Appendix A.1.1 - log-likelihood reward**, we first prove the loss decomposition when the reward is the reconstruction log-likelihood (case of the main paper).
  - **Appendix A.1.2 - general reward**, we then extend the decomposition to a more general reward
- **Appendix A.2 - Extension to Lewis games**, we extend the Loss Decomposition to a more general class of Lewis games. We first describe the additional formalism (Appendix A.2.1), then we prove the decomposition when the reward is the listener's log-likelihood (Appendix A.2.2) and when the reward is more general (Appendix A.2.3). Eventually, we show how the classic discrimination game can be expressed under this formalism in Appendix A.2.4.
- **Appendix A.3 - Extension to agents optimizing different rewards**, we discuss how the decomposition is affected when the agents optimize different rewards.

A.1 Proof of the Lewis Reconstruction Game Loss Decompositon

Let's first recall some notations that we will use throughout the proofs. We consider two agents: a speaker parameterized by θ and a listener parameterized by ϕ. In the reconstruction game, the speaker observes objects denoted by x and taken from a set $\mathcal{X}$. The random variable characterizing the object is denoted by X and its distribution is denoted by p. Based on object x, the speaker then sends a message m from a message space $\mathcal{M}$ according to its policy $\pi_\theta(.|x)$. The random variable M_θ characterizes the message that is sampled from the speaker's policy π_θ. Eventually, the listener should reconstruct the original object x based on the message m. The probability that the listener predicts the input x given a message m is denoted by $\rho_\phi(x|m)$.

For any probability distribution, we denote by Supp the support of the distribution.

In the reconstruction game, the two agents optimize the same loss:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[r_\phi(x,m)]$$

We will first prove the decomposition in the case where $r_\phi(x,m) = \log \rho_\phi(x|m)$ (reconstruction log-likelihood) for all x and m and then for a more general form of reward.

A.1.1 Proof of the Decomposition when $r_\phi(x,m) = \log \rho_\phi(x|m)$

We first prove the decomposition in the case described in the main paper: $r_\phi(x,m) = \log \rho_\phi(x|m)$ for all x and m.

Optimal listener For completeness, we recall the proof of Equation 2 of the expression of the listener that is optimal with respect to $\mathcal{L}_{\theta,\phi}$.

In the case $r_\phi(x,m) = \log \rho_\phi(x|m)$, the listener is optimizing a cross-entropy loss with respect to the joint variable (X, M) where X follows p and M follows speaker's policy π_θ. The loss can be rewritten as:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho_\phi(x|m)]$$
$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^\theta(.|m)}[\log \rho_\phi(x|m)]$$

According to Gibbs inequality, the optimal distribution $\rho_{\phi^*}(\cdot|m)$ for all m is $\rho_{\phi^*}(.|m) = \rho^{*(\theta)}(.|m)$ where $\rho^{*(\theta)}(.|m)$ is the speaker's posterior distribution with respect to the prior p and the conditional distribution $\pi_\theta(\cdot|x)$:

$$\rho^{*(\theta)}(x|m) := \frac{p(x)\pi_\theta(m|x)}{\mathbb{E}_{x'\sim p}[p(x')\pi_\theta(m|x')]} \qquad \text{for all} \quad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(\cdot|x))$$

This concludes the proof of Equation 2.

Loss Decomposition The idea of the proof is to decompose the reward into the optimal reward (when the listener is optimal), denoted by $r^{*(\theta)}(x, m)$, and the residual that measures the optimality gap, denoted by $r_\phi(x, m) - r^\theta(x, m)$:

$$r_\phi(x, m) = r^{*(\theta)}(x, m) + (r_\phi(x, m) - r^{*(\theta)}(x, m)) \qquad \text{for all} \qquad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(.|x))$$

Due to the linearity of the expectation, it follows that:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)]$$

In the case where the reward is taken as the listener's log-likelihood, we have:

$$\begin{aligned}
\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{*(\theta)}(x|m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}\left[\log \frac{\rho_\phi(x|m)}{\rho^{*(\theta)}(x|m)}\right] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{*(\theta)}(x|m)] - \mathbb{E}_{m\sim\pi_\theta}\mathbb{E}_{x\sim\rho^{*(\theta)}(.|m)}\left[\log \frac{\rho_\phi(x|m)}{\rho^{*(\theta)}(x|m)}\right] \\
\mathcal{L}_{\theta,\phi} &= \underbrace{\mathcal{H}(X|M_\theta)}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi_\theta} D_{KL}(\rho^{*(\theta)}(\cdot|m)||\rho_\phi(\cdot|m))}_{\mathcal{L}_{\text{adapt}}}
\end{aligned}$$

where $\mathcal{H}(X|M_\theta)$ is the conditional entropy of X conditioned on M_θ and $D_{KL}(p||q)$ is the Kullback-Leiber divergence between two distributions p and q.

This last computation concludes the proof of Equation 5.

Remarks The key ingredients of the loss decomposition are:

1. We isolate two sub losses: $\mathcal{L}_{\text{info}}$, independent from the listener ; $\mathcal{L}_{\text{adapt}}$ optimized both by the speaker and the listener.
2. $\mathcal{L}_{\text{info}}$ measures the degree of ambiguity in the communication protocol. If $\mathcal{L}_{\text{info}}$ is optimal, ie. $\mathcal{L}_{\text{info}} = 0$, messages are unambiguous: each message refers to a unique input. Otherwise, $\mathcal{L}_{\text{info}} > 0$ and ambiguities remain.
3. $\mathcal{L}_{\text{adapt}}$ measures the gap between the listener and its optimum (here the speaker's posterior distribution). When the listener is optimal, $\mathcal{L}_{\text{adapt}} = 0$ and the main loss is limited to its information part, otherwise $\mathcal{L}_{\text{adapt}} > 0$ and the speaker and listener should adapt to reduce the optimality gap.

A.1.2 Decomposition with a General Reward

In order to generalize the loss decomposition to more general rewards, we adopt the following strategy:

- **Construction of the reward**: we first need to build a general expression of the communication reward. To do so, we describe the conditions that the cooperative reward should fulfill in the reconstruction game and then propose a general reward expression. For the sake of generality, we consider that the environment $\mathcal{X}$ and message space $\mathcal{M}$ may be continuous spaces and that all the probability distribution may not be discrete.

- **Examples of usual cases**: we show that our proposed general expression covers the rewards used in most emergent communication papers, e.g. log-likelihood and accuracy.
- **Loss decomposition in the general case**: we write the loss decomposition with this general form of reward, showing that the key properties of the loss decomposition still hold.

Construction of the reward The Lewis reconstruction game is a cooperative game: the more the listener is able to reconstruct the objects seen by the speaker, the better the task is solved both by the speaker and the listener. Therefore the reward of the Lewis reconstruction game should respect the following conditions:

- **C1**: For $x \in \text{Supp}(p)$ and $m \in \text{Supp}(\pi_\theta(.|x))$, the expected reward $r_\phi(x, m)$ is maximum when $\rho_\phi(.|m) = \mathbf{1}_x$, where ie. $\mathbf{1}_x$ denotes the indicator function on $\mathcal{X}$ taken on x: the listener predicts x with probability 1 when it receives m.
- **C2**: For $x \in \text{Supp}(p)$ and $m \in \text{Supp}(\pi_\theta(.|x))$, the expected reward $r_\phi(x, m)$ is sub-optimal when $\rho_\phi(.|m) \neq \mathbf{1}_x$, ie., the listener has a non-negative probability to predict the wrong object $x' \neq x$.

Given these assumptions, we propose the following general reward expression:

$$r_\phi(x, m) = -D(\mathbf{1}_x || \rho_\phi(.|m)) + K \tag{11}$$

where D is such that $D(p||q) = 0$ iff $p = q$,$D(p||q) > 0$ otherwise. $\mathbf{1}_x$ is the indicator function on $\mathcal{X}$ taken in x and K is a real number that fixes the highest value of the reward. Note that $D(p||q)$ is close to a divergence, but has less assumptions.

Usual rewards as special instances of the general expression

We show that Equation 11 recovers most rewards used in the emergent communication literature, specifically:

- **Reconstruction log-likelihood** [9, 10, 40, 68, 67, 11] This case is the one used in the main paper and which is standarly used in reconstruction settings. With the following parameters
  - $D(p||q) = D_{KL}(p||q)$,
  - $K = 0\cdot$

  we have:

  $$r_\phi(x, m) = -D_{KL}(\mathbf{1}_x || \rho_\phi(.|m)) = \log \rho_\phi(x|m)$$

- **Accuracy** [65, 53, 52, 48, 56, 28, 23] Accuracy is the most commonly used reward in the emergent communication literature. It corresponds to agents receiving reward 1 if the prediction sampled according to the listener's output probability $\rho_\phi(\cdot|m)$ matches the original object x. The pointwise accuracy depends on the specific sample drawn from the listener's distribution. We are interested in how good the listener's prediction is on average, and thus in the expected accuracy, which is expressed as in Equation 11 with:
  - $D(p||q) = 1 - \mathbb{E}_p[q]$
  - $K = 1$

  The expected accuracy is then defined as:

  $$r_\phi(x, m) = D(\mathbf{1}_x || \rho_\phi(.|m)) = \rho_\phi(x|m)\cdot$$

Loss Decomposition

With this definition of the reward, the speaker and listener loss can be written:

$$\begin{aligned}\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[r_\phi(x|m)] \\ &= -K + \mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_x || \rho_\phi(.|m))]\end{aligned}$$

We first need to define the optimal listener. Note that the expectation can be re-formulated:

$$\mathcal{L}_{\theta,\phi} = -K + \mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(\cdot|m)}[D(\mathbf{1}_x||\rho_\phi(.|m))]$$

The optimal listener is the listener ρ_ϕ that minimizes $\mathbb{E}_{x\sim\rho^{*(\theta)}(\cdot|m)}[D(\mathbf{1}_x, \rho_\phi(.|m))]$ for all m. In the general case, there is no close-formed expression of the optimal listener. The optimal listener policy is dependent of the function D and the posterior distribution $\rho^{*(\theta)}(\cdot|m)$. We next denote the optimal listener policy $\rho_{\phi^*}(\cdot|m)$ that is fully characterized by $\rho^{*(\theta)}(\cdot|m)$ and D and is **independent** of ϕ.

As in Appendix A.1.1, we denote $r^{*(\theta)}(x, m)$ the reward of the optimal listener. We can then apply the same reward decomposition as in Appendix A.1.1:

$$r_\phi(x, m) = r^{*(\theta)}(x, m) + (r_\phi(x, m) - r^{*(\theta)}(x, m)) \qquad \text{for all} \qquad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(.|x))$$

which is equal to:

$$r_\phi(x, m) = -D(\mathbf{1}_x||\rho_{\phi^*}(.|m)) - [D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(.|m))] - K$$

where $\rho_{\phi^*}(.|m)$ is the optimal listener distribution that is independent of ϕ.

We can then rewrite the loss by taking the expectation of this reward and isolate an information and co-adaptation component:

$$\begin{aligned}
\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)] \\
&= \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{info}}} \\
&+ \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{adapt}}} - K
\end{aligned}$$

To be an information/co-adaptation decomposition, this loss decomposition should fulfill the following conditions:

1. $\mathcal{L}_{\text{info}}$ should be independent from the listener's weight ϕ; $\mathcal{L}_{\text{adapt}}$ should be optimized both by the speaker and the listener.
2. $\mathcal{L}_{\text{info}}$ should be optimal ($\mathcal{L}_{\text{info}} = 0$) when the communication protocol is unambiguous, ie. each message refers to a unique input, sub-optimal ($\mathcal{L}_{\text{info}} > 0$) otherwise.
3. $\mathcal{L}_{\text{adapt}}$ should be 0 when the listener matches its optimum value with respect to the current object-message joint distribution, otherwise $\mathcal{L}_{\text{info}} > 0$ and the speaker and listener should adapt to reduce the optimality gap.

Let's prove that all those conditions hold:

1. The optimal listener policy $\rho_{\phi^*}(\cdot|m)$ is independent of ϕ. It turns out that $\mathcal{L}_{\text{info}}$ is independent from the listener. On the contrary, $\mathcal{L}_{\text{adapt}}$ is dependent both on θ and ϕ and therefore is optimized both by the speaker and listener.
2. Let first show that when $\mathcal{L}_{\text{info}} = 0$, the speaker language is unambiguous. The language is considered unambiguous iff each message refers to a unique input. Formally, let x be in the support of p and

   $$\mathcal{M}_x = \{m \in \mathcal{M} \quad | \quad \rho^{*(\theta)}(x|m) > 0\},$$

   be the set of messages referring to x.

This set is non empty because $\mathbb{E}_{m\sim\pi_\theta}[\rho^{*(\theta)}(x|m)] = p(x) > 0$. The emergent language is considered unambiguous iff for all x and x' in the support of p:

$$x \neq x' \Rightarrow \mathcal{M}_x \cap \mathcal{M}_{x'} = \emptyset,$$

This property is equivalent of having a speaker posterior distribution $\rho^{*(\theta)}(.|m)$ being a Dirac distribution for all m (otherwise, there is at least one message that refers to more than one object).

Let's demonstrate that $\mathcal{L}_{\text{info}} = 0$ iff $\rho^{*(\theta)}(.|m)$ is a Dirac distribution for all m.

First, when the speaker's posterior distribution is not a Dirac distribution, we have: $\mathcal{L}_{\text{info}} > 0$. Let m be a message in the support of π_θ. If $\rho^{*(\theta)}(\cdot|m)$ is not a Dirac distribution, there exists x such that $\rho^{*(\theta)}(x|m) > 0$ and $D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m)) > 0$. Indeed, if there exists x' such that $D(\mathbf{1}_{x'}||\rho_{\phi^*}(\cdot|m)) = 0$, we have $\rho_{\phi^*}(\cdot|m) = \mathbf{1}_{x'}$ by definition of D and thus: if $x \neq x' \quad \Rightarrow \quad D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m)) = D(\mathbf{1}_x||\mathbf{1}_{x'}) > 0$ by definition of D. It implies that when $\rho^{*(\theta)}(\cdot|m)$ is not a Dirac distribution : $\mathbb{E}_{x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))] > 0$.

Reciprocally, if for all $m \in \text{Supp}(\pi_\theta)$, $\rho^{*(\theta)}(\cdot|m)$ is a Dirac distribution: $\rho^{*(\theta)}(\cdot|m) = \mathbf{1}_{x_m}$ (with m referring to x_m and all $x \in \text{Supp}(p)$ covered by the messages) and the corresponding optimal listener is also the Dirac distribution $\rho_{\phi^*}(\cdot|m) = \mathbf{1}_{x_m}$, we have:

$$\mathcal{L}_{\text{info}} = \mathbb{E}_{m\sim\pi_\theta}[D(\mathbf{1}_{x_m}||\rho_{\phi^*}(\cdot|m))] = \mathbb{E}_{m\sim\pi_\theta}[D(\mathbf{1}_{x_m}||\mathbf{1}_{x_m})] = 0$$

Therefore, $\mathcal{L}_{\text{info}}$ is equal to 0, ie. is minimum, if and only if the speaker has a posterior which is Dirac distribution, ie. the speaker develops an unambiguous language.

3. When the listener is optimal with respect to its loss, $\rho_\phi(.|m) = \rho_{\phi^*}(.|m)$ for all m and as a direct consequence, $\mathcal{L}_{\text{adapt}} = 0$. When the listener is not optimal with respect to its loss, $\mathcal{L}_{\text{adapt}} > 0$ by definition of the optimal listener which is the listener that minimizes $\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^\theta(\cdot|m)}[D(\mathbf{1}_x||\rho_\phi(.|m))]$.

In conclusion, in the case of a general reward, we keep the main ingredients of the information/co-adaptation decomposition.

A.2 General Proof of the Lewis Games Loss Decomposition

In the previous section, we provided a proof of the loss decomposition for the Lewis Reconstruction Game with a general cooperative reward. The goal of this Section is to extend this decomposition to a more general definition of Lewis Games:

- **Appendix A.2.1 - Formalism**: We first describe the additional formalism.
- **Appendix A.2.2 - Log-likelihood reward**: We prove the decomposition for the general Lewis Game when the reward is the listener's log-likelihood.
- **Appendix A.2.4 - General cooperative reward** We prove the decomposition for the general Lewis Game with a general cooperative reward.
- **Appendix A.2.4 - Discrimination game** : Eventually, we show how the widely studied discrimination game [12, 60, 21, 30, 65, 53, 52, 33, 56, 58] can be expressed under this formalism.

A.2.1 Formalism

In the general form, we consider inputs x from a set $\mathcal{X}$ where x is drawn from p_X. We consider a random feature F of X (in the reconstruction game $F = X$) that is distributed following $p_F(.|X)$. A draw of F is denoted f and the set of potential features $\mathcal{F}$. We here consider that the listener may have access to an auxiliary input y. We denote Y the random variable of this auxiliary input and $p_Y(\cdot|X, F)$ its probability distribution. The task is here the communication of the feature f. To this end, the speaker still sends messages m from the message space $\mathcal{M}$. The random variable M_θ characterizes the messages that are sampled from the speaker's policy $\pi_\theta(\cdot|X)$. Eventually, the probability that the listener predicts the correct feature f, given message m and auxiliary features y is denoted by $\rho_\phi(f|m, y)$.

A.2.2 Proof with the log-likelihood reward: $r_\phi(f, m, y) = \log \rho_\phi(f|m, y)$

We first prove the decomposition in the case: $r_\phi(f, m, y) = \log \rho_\phi(f|m, y)$ for all f,m and y, ie. the reward is the listener's log-likelihood of predicting the good feature. The agents' loss becomes

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(f|m, y)].$$

Optimal listener The optimal listener is the listener that optimally minimizes $L_{\theta,\phi}$ for a fixed speaker policy π_θ. It is obtained by noting that:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(f|m, y)]$$
$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{(m,y)\sim p_{M_\theta,Y}} \mathbb{E}_{f\sim\rho^\theta(\cdot|m,y)}[\log \rho_\phi(f|m, y)].$$

where $\rho^{*(\theta)}(f|m, y) = \frac{\mathbb{E}_{x\sim p_X}[\pi_\theta(m|x)p_F(f|x)p_Y(y|f,x)]}{\mathbb{E}_{x\sim p_X, f\sim p_F(.|x)}[\pi_\theta(m|x)p_Y(y|f,x)]}$ for all f, m and y and $p_{M_\theta,Y}(m, y) = \mathbb{E}_{x\sim p_X, f\sim p_F(.|x)}[\pi_\theta(m|x)p_Y(y|f, x)]$.

It follows from Gibbs inequality that the optimal listener is $\rho^{*(\theta)}(\cdot|m, y)$ for all m and y.

We can apply the reward decomposition of Appendix A.1.1:

$$r_\phi(x, m) = r^{*(\theta)}(f, m, y) + (r_\phi(f, m, y) - r^{*(\theta)}(f, m, y)) \qquad \text{for all} \qquad f \in \mathcal{F}, m \in \mathcal{M}, y \in \mathcal{Y}.$$

Plugging this decomposed reward in our loss, and applying the exact same steps as in Appendix A.1.1, we get

$$\mathcal{L}_{\theta,\phi} = \underbrace{\mathcal{H}(F|M_\theta, Y)}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{(m,y)\sim p_{M_\theta,Y}} D_{KL}(\rho^{*(\theta)}(\cdot|m, y)||\rho_\phi(\cdot|m, y))}_{\mathcal{L}_{\text{adapt}}} \tag{12}$$

which is the Loss Decomposition for a general game.

Remarks You note that the decomposition is close to the Loss Decomposition in the reconstruction case (Equation 5). Indeed, since the listener should predict a given feature F, the information task is to build an unambiguous message protocol with respect to this feature and the optimal listener becomes the posterior distribution of the speaker with respect to this feature. The co-adaptation loss is once again a Kullback-Leiber distribution between the listener and the speaker's posterior. $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$ respects the conditions states in Appendix A.1.1.

A.2.3 Proof with the general reward $r_\phi(f, m, y) = -D(\mathbf{1}_f||\rho_\phi(.|m, y)) + K$

To study the general case, we use the reward definition provided in Appendix A.1.2:

$$r_\phi(f, m, y) = -D(\mathbf{1}_f||\rho_\phi(.|m, y)) + K$$

where $D(p||q)$ is a function that is null when $p = q$, greater than 0 otherwise, $\mathbf{1}_f$ the indicator function on $\mathcal{F}$ taken in f and K is a real number that fixes the highest value of the reward.

Agents' loss becomes:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r_\phi(f, m, y)]$$
$$\mathcal{L}_{\theta,\phi} = \mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_\phi(.|m, y))] - K$$

Denoting $\rho_{\phi^*}(.|m, y)$ the listener that optimally minimises $\mathcal{L}_{\theta,\phi}$ and $r^\theta(f, m, y)$ the reward of the optimal listener, the loss can be decomposed:

$$
\begin{aligned}
\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r^{*(\theta)}(f,m,y)] \\
&\quad - \mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r_\phi(f,m,y) - r^{*(\theta)}(f,m,y)] \\
\mathcal{L}_{\theta,\phi} &= \underbrace{\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_{\phi^*}(.|m,y))]}_{\mathcal{L}_{\text{info}}} \\
&\quad + \underbrace{\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_\phi(.|m,y)) - D(\mathbf{1}_f||\rho_{\phi^*}(.|m,y))]}_{\mathcal{L}_{\text{adapt}}} - K
\end{aligned}
$$

For the same arguments as in Appendix A.1.2, $\mathcal{L}_{\text{info}}$ is only optimized by the speaker and is optimal when the speaker develops an unambiguous message protocol with respect to F given Y, $\mathcal{L}_{\text{adapt}}$ is null when the listener is optimal, otherwise it is > 0, ie. sub-optimal. Therefore, we recover the key ingredients of the Loss Decomposition: when the listener is optimal, speaker's loss is limited to $\mathcal{L}_{\text{info}}$, when the listener is not optimal, the speaker has the additional task to help the listener matching its optimum.

A.2.4 Case of the Discrimination Game

Recall that in a discrimination game, as in a reconstruction game, the speaker observes an input, x and sends a message m to the listener. The listener is then provided with both the message m, and a list of $N+1$ candidate inputs, containing input x, along with N other inputs, or *distractors*. The goal of the listener is then to give the index of the candidate that corresponds to the actual input.

To formally define discrimination games as instances of the general Lewis game described above, we define $X_1, \ldots, X_N$ to be i.i.d. samples from the inputs distribution p. These inputs will be used as the distractors. We additionally set $X_0 = X$. We then define a random permutation Σ, drawn uniformly from the set of $N+1$ element permutations, and independently from all other random variables. We then set our auxiliary input $Y = (X_{\Sigma(0)}, \ldots, X_{\Sigma(N)})$, which provides the listener with a permuted list, containing both the correct input at a random position, as well as the distractors. Finally, we set the feature to be predicted as $F = \Sigma^{-1}(0)$. The task of the listener becomes to identify the index of the correct input among all distractors, and we recover a discrimination game.

A.3 Speaker and Listener Optimizing Different Rewards

In this paper, we only discuss the case where the agents are fully cooperative, ie. they are optimizing exactly the same reward. When the agents are not aligned on the same objective, the system should be decoupled and an additional *alignement bias* is added to the loss of the speaker. For example, in the reconstruction game where the speaker is optimizing a general reward $r_\phi(x,m) = -D(\mathbf{1}_x, \rho_\phi(\cdot|m)) + K$ and the listener a cross-entropy loss, the system becomes:

$$
\begin{cases}
\mathcal{L}_\theta &= \mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_x, \rho_\phi(\cdot|m))] - K \\
\mathcal{L}_\phi &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)].
\end{cases}
\tag{13}
$$

where $\mathcal{L}_\theta$ is the speaker's loss and $\mathcal{L}_\phi$ the listener's loss.

By denoting $\rho_{\phi^*}(\cdot|m)$ the optimal listener for all m with respect to $\mathcal{L}_\theta$ (which is fully determined by the speaker's posterior and D) and $\rho^{*(\theta)}(\cdot|m)$ the optimal listener for all m with respect to $\mathcal{L}_\phi$ (in this case, the speaker posterior), the speaker loss now decomposes into:

$$
\begin{aligned}
\mathcal{L}_\theta &= \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho^{*(\theta)}(\cdot|m))]}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{adapt}}} \\
&\quad + \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(.|m)) - D(\mathbf{1}_x||\rho^{*(\theta)}(\cdot|m))]}_{\text{alignment bias}} - K
\end{aligned}
$$

Compared to the standard decomposition, there is an additional term, that we name the *alignment bias*, linked to the gap between the listener optimum of $\mathcal{L}_\theta$ and the listener optimum of $\mathcal{L}_\phi$. If those

optima are close, the amplitude of this term is negligible compared to $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$. If those optima are very different (eg. competitive game), the information and co-adaptation terms could have a significantly smaller amplitude compared to the *alignment bias*. We leave to future work the theoretical study of this *alignment bias* which echoes some empirical studies [61].

B Method: Additional Computations

In Section 3.2, we propose a protocol to balance the importance of the information and co-adaptation losses in the speaker's training loss. To do so, we use the probe listener's estimate of the speaker's posterior on the train set $\rho^{\text{train}}_{\omega^*}(x|m) = \log \rho^{\text{train}}_{\omega^*}(x|m)$ and build the following reward:

$$r_\phi(x, m; \alpha) = (1 - 2\alpha) \times \underbrace{\log \rho^{\text{train}}_{\omega^*}(x|m)}_{\text{probe listener reward}} + \alpha \times \underbrace{\log \rho_\phi(x|m)}_{\text{standard listener reward}}$$

where α is a weight in $[0; 0.5]$.

The loss equality defined in Section 3.2 is then recovered with the following computations:

$$\begin{aligned}
\mathcal{L}_\theta(\alpha) &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m; \alpha)] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[(1 - 2\alpha) \times \log \rho^{\text{train}}_{\omega^*}(x|m) \ + \ \alpha \times \log \rho_\phi(x|m)] \\
&= -(1 - 2\alpha)\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{\text{train}}_{\omega^*}(x|m)] - \alpha\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho_\phi(x|m)] \\
&= (1 - 2\alpha)\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \alpha(\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}) \\
\mathcal{L}_\theta(\alpha) &= (1 - \alpha)\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \alpha\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}
\end{aligned}$$

Remark In the paper, we only consider the case $\alpha \in [0; 0.5]$ and do not explore larger values of α. Indeed, controlling the co-adaptation rate α is made by re-weighting $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ (estimated with a probe listener). However, two issues occur when $\alpha > 0.5$:

- First, the goal of computing $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is to indirectly balance the weight of the training information loss $\mathcal{L}^{train}_{\text{info}}$. By taking the loss of the probe listener close to optimality, we get an upper bound estimate $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ of the training information loss $\mathcal{L}^{\text{train}}_{\text{info}}$. Therefore, it theoretically ensures that we minimize $\mathcal{L}^{\text{train}}_{\text{info}}$ when optimizing $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$. However, when $\alpha > 0.5$, the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is negative. In this case, since $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is an upper bound of $\mathcal{L}^{\text{train}}_{\text{info}}$, we do not have the guarantee that the speaker minimizes $-\mathcal{L}^{\text{train}}_{\text{info}}$ anymore.
- Second, we empirically experimented $\alpha > 0.5$ even if theoretical conditions are not reached. In practice, if the system converged for values of α closed to 0.5, the system quickly became unstable for larger values of α. Our main hypothesis is that the speaker cannot start structuring its messages when the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}$ is too strong. Indeed, agents start with random weights. It implies that, at the beginning of the training, if the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}$ is too strong, it pressures the speaker to have an almost uniform posterior, ie. to develop a fully ambiguous language. In short, if α is too large, the speaker has too little pressure on developing meaningful messages and therefore succeeding in the communication task.

C Regularization

We here provide:

- the parameters used for the listener's regularization (Appendix C.1)
- the results obtained when regularizing the speaker (Appendix C.2)

C.1 Parameters of the Listener's Regularization

Regularization parameters have been tuned in order to get the best average generalization scores while having a convergence success rate greater or equal to 75%. When regularizing with the layer

normalization (noted *No LN.* in Table 1), we remove the layer normalization applied of the listener's LSTM cell. Dropout rate is set to 0.2 and weight decay penalty is set to 0.01 both when layer normalization is kept (noted *Weight decay* in Table 1) and when layer normalization is removed (noted *No LN.* + *WD* in Table 1).

C.2 Comparison with Speaker's Regularization

Parameters For the sake of completeness, we also study the impact of regularizing the speaker. Here, we only report the results with the weight decay penalty. Indeed, removing the layer normalization makes the training slow and unstable while results with dropout are worse than those with weight decay. Weight decay penalty has been fine-tuned to 0.005 to get the best average generalization performances while having $> 75\%$ successful experiments.

Results In Table 2, we compare the generalization and compositionality of emergent languages with and without regularization applied on the speaker. First, when we regularize the speaker without any regularization on the listener, we see that the gain of generalization and compositionality is negligible and inferior to the gain obtained when regularizing the listener. Moreover, we note that when we regularize both the speaker and the listener, scores of generalization and compositionality are similar to those obtained when only regularizing the listener. It suggests that regularizing the speaker has little impact on generalization and compositionality.

These results support the claim of Section 5.2: the listener is the main contributor of the co-adaptation overfitting in the reconstruction game.

No Speaker reg.	Gen. ↑	Compo. ↑
Continuous	$0.58_{\pm 0.05}$	$0.22_{\pm 0.02}$
No LN.	$0.70_{\pm 0.03}$	$0.24_{\pm 0.02}$
Weight decay	$0.72_{\pm 0.03}$	$0.25_{\pm 0.03}$
No LN. + WD	$0.87_{\pm 0.07}$	$0.30_{\pm 0.03}$

Speaker with WD	Gen. ↑	Compo. ↑
Continuous	$0.62_{\pm 0.02}$	$0.22_{\pm 0}.03$
No LN.	$0.68_{\pm 0.07}$	$0.23_{\pm 0.01}$
Weight decay	$0.74_{\pm 0.05}$	$0.26_{\pm 0.04}$
No LN. + WD	$0.82_{\pm 0.07}$	$0.32_{\pm 0.04}$

Table 2: Performance comparison: (left) without speaker regularization ; (right) with speaker regularization. Weight decay penalty on the speaker is set to 0.005. Parameters of regularization methods for the listener are reported in Appendix C.1.

D Image Discrimination Games

We here complete Section 5.3 by presenting the rules and experimental settings of the image discrimination game (Appendix D.1), reporting the results of compositionality (Appendix D.2) and completing generalization results of Table 1 with regularization experiments (Appendix D.3).

D.1 Experimental Settings

For the implementation of the image discrimination game, we mostly follow the protocol proposed by [12].

D.1.1 Game Rules and notations

In the Lewis image discrimination game, the speaker observes an image. Then, the speaker sends a descriptive message to the listener. Based on this message, the listener should retrieve the correct image among a set of candidates.

Formally, the image observed by the speaker is denoted by x and belongs to a set $\mathcal{X}$. The intermediate message sent by the speaker is denoted by m and belongs to a set a potential messages $\mathcal{M}$. The speaker follows a policy π_θ which samples a message m with probability $\pi_\theta(m|x)$ conditioned on image x. The listener encodes the message m into a representation $t_\phi(m)$. The set of candidates received by the listener are denoted $\mathcal{C}$ and the listener encodes each candidates $x' \in \mathcal{C}$ by a representation $t_\phi(x')$. The probability of a candidate x' to be the correct image is : $\rho_\phi(x'|m, \mathcal{C})$. It is obtained by comparing the message encoding $t_\phi(m)$ with the image encoding $t_\phi(x')$ of all candidates.

D.1.2 Environment

Datasets We perform the discrimination game on ImageNet [19, 69] and CelebA [57]. We work with image pre-processed encodings $f(x)$ of size 2048 that have been open-sourced by [12]. In the two datasets, each image has been center-cropped and processed by a ResNet-50 encoder pretrained on ImageNet with the self-supervised method BYOL [29].

Train/val/test splits For building our custom training sets, we first considered the splits provided by [12]. From the respective $1400k$ and $200k$ labelled images of ImageNet and CelebA, they slitted the dataset in train, validation and test with the ratio $80/10/10$.

To test agents generalization capacities, we also build subsets of the training set provided by [12]: ImageNet $\frac{1}{20}$, ImageNet $\frac{1}{100}$, CelebA $\frac{1}{20}$ and CelebA $\frac{1}{100}$. For each of those sub-training sets, we randomly selected a small fraction of the training set, approximatively corresponding to $1/20$-th and $1/100$-th of the total training set. The corresponding number of samples are reported in Table 3.

	Training samples			Training samples	
CelebA	$1/20$	$1/100$	**ImageNet**	$1/20$	$1/100$
	8492	2123		50732	12683

Table 3: Number of training samples for the four training subsets considered: ImageNet $\frac{1}{20}$, ImageNet $\frac{1}{100}$, CelebA $\frac{1}{20}$ and CelebA $\frac{1}{100}$

All our experiments on images are run with those 4 small training sets. We keep the original validation and test sets from [12].

D.1.3 Agent Models

Speaker model The speaker is a neural network that takes the pre-processed representation of an image $f(x)$ as input of size 2048 and returns a message $m = (m_i)_{1 \leq i \leq T}$ of length T.

The speaker follows a recurrent policy: given the image representation $f(x)$, it samples for all $t \in [1, T]$ a token m_t with probability $\pi_\theta(m_t | m_{<t}, f(x))$. The image representation $f(x)$ is first projected by a linear layer to get an object embedding of size 256 that is used to initialize a LSTM of size 256 with layer normalization. At each time step, the LSTM's output is fed into a linear layer of size $|\mathcal{V}|$, followed by a softmax, to produce $\pi_\theta(m_t | m_{<t}, f(x))$.

In our experiments, the following parameters have been chosen: $T = 10$, $|V| = 10$ meaning that the message space is of size 10^{10} preventing any channel capacity bottleneck.

Listener model The listener is a neural network that takes the speaker's message m and a set of image candidates $\mathcal{C}$ containing the target image x and outputs the probability for each candidate $x' \in \mathcal{C}$ to be the target image x.

The listener is composed of two modules: one that encodes the message ; the other that encodes images. For a message $m = (m_1, ..., m_T)$, the listener passes each symbol m_t through an embedding layer of dimension 256 followed by a LSTM of size 256 with layer normalization. The final recurrent state h_T^1 is then passed to a linear layer that produces the image encoding $t_\phi(m)$ of size 256. In parallel, each candidate x' is first pre-processed by f and then passed through a linear layer producing an image encoding $t_\phi(x')$ of size 256.

The message representation $t_\phi(m)$ is then compared to each candidate representation $t_\phi(x')$ with the following score function: $\text{score}(m, x', \phi) := t_\phi(m) t_\phi(x')^T$. Note that contrary to [12], we rather use a dot-product score function [53] instead of a cosine similarity because we empirically got better results and more stable trainings. The probability distribution over the candidates $\mathcal{C}$ of being the target image x is then obtained by normalizing the scores with a softmax. This probability distribution is denoted by $\rho_\phi(\cdot | m, \mathcal{C})$ and the listener guess is $\hat{x} = \underset{x'}{\text{argmax}} \rho_\phi(x' | m, \mathcal{C})$.

In our experiments, the number of candidates is $|\mathcal{C}| = 1000$.

D.1.4 Agents Training

We follow the same principle as in the reconstruction game: the listener is trained to best predict the target image among the set of candidates, while the speaker takes the opposite of the listener's loss as reward:

Listener loss The listener is trained to predict the target image among the set of candidates $\mathcal{C}$. When receiving a batch of inputs x, a set of candidates $\mathcal{C}$ is sampled for each input x. The sampling is uniform without replacement over $\mathcal{X} - \{x\}$ meaning that the target image x cannot be duplicated into the candidates. The listener is then trained to optimized the average InfoNCE loss [62]:

$$\mathcal{L}_\phi = \sum_{x \in \text{batch}} -\log \rho_\phi(x|m, \mathcal{C})$$

Speaker loss When the speaker observes an image x, sends a message m and the listener has to choose among a set of candidates $\mathcal{C}$, the speaker's reward is defined as:

$$r_{\phi,\mathcal{C}}(x, m) = \log \rho_\phi(x|m, \mathcal{C})$$

The speaker is trained to maximize its cumulative reward: $\mathbb{E}_{x,m,\mathcal{C}}[\log \rho_\phi(x|m, \mathcal{C})]$ which means that the speaker and the listener have the same loss.

Optimization The agents are optimized using `Adam` [42] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The speaker's learning rate is $5 \cdot 10^{-4}$ while the listener's learning rate is $1 \cdot 10^{-3}$. Agents are trained on batches of size of 2048. For the speaker, we use policy gradient [76], with a baseline computed as the average reward within the minibatch, and we add an entropy regularization of 0.01 to the speaker's loss [82].

D.2 Topographic Similarity Results

We report results of topographic similarity for experiments of Section 5.3. To be complete, we add the scores when applying listener regularization (corresponding generalization performances are reported in Appendix D.3).

Scores of topographic similarity are reported in Table 4. Here, the distance used to compare images is the cosine distance between the vector representations of the ResNet-50 encoder pretrained on ImageNet. The distance used to compare messages remains the edit-distance. As mentioned in the main paper, we can see that there is not any compositionality trend when agents communicate about images. Moreover, when comparing with Table 6 that reports generalization performances, we see that gains of generalization do not correlate with gains of topographic similarity. It suggests that the topographic similarity does not capture agents' language structure in image based settings, as already observed in previous work [12, 1].

	Topographic similarity ↑			Topographic similarity ↑	
CelebA	1/20	1/100	**ImageNet**	1/20	1/100
Continuous	$\mathbf{0.28}_{\pm \mathbf{0.03}}$	$\mathbf{0.32}_{\pm \mathbf{0.03}}$	Continuous	$0.17_{\pm 0.03}$	$0.17_{\pm 0.03}$
No LN.	$0.26_{\pm 0.04}$	$0.29_{\pm 0.03}$	No LN.	$0.18_{\pm 0.01}$	$0.16_{\pm 0.02}$
No LN. + WD	–	$0.30_{\pm 0.03}$	No LN. + WD	$\mathbf{0.19}_{\pm \mathbf{0.03}}$	$\mathbf{0.21}_{\pm \mathbf{0.03}}$
Weight decay	$0.27_{\pm 0.03}$	$0.28_{\pm 0.04}$	Weight decay	$0.17_{\pm 0.02}$	$0.15_{\pm 0.02}$
Early stopping	$0.27_{\pm 0.04}$	$0.30_{\pm 0.03}$	Early stopping	$0.18_{\pm 0.04}$	$0.20_{\pm 0.03}$

Table 4: Topographic similarity of emergent languages in the image discrimination game where images are compared with a cosine similarity. *No LN.* refers to the removal of the layernorm on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to 0.01 ; *No LN. + WD* refers to the removal of layernorm and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.

In addition, we also test whether scores of topographic similarities are improved when using another distance to compare images. In Table 5, we use the attributes provided in CelebA to compare the

images. The distance between two images is computed as $1 -$ (propotion of common attributes). For the message comparison, we keep the edit-distance. Once again, no topographic similarity trends emerge, sustaining results already observed in [12].

	Topographic similarity (with attributes) ↑	
CelebA	$1/20$	$1/100$
Continuous	$0.13_{\pm 0.02}$	$\mathbf{0.15_{\pm 0.03}}$
No LN.	$\mathbf{0.14_{\pm 0.02}}$	$0.14_{\pm 0.03}$
No LN. + WD	–	$\mathbf{0.15_{\pm 0.04}}$
Weight decay	$\mathbf{0.14_{\pm 0.02}}$	$\mathbf{0.15_{\pm 0.01}}$
Early stopping	$0.13_{\pm 0.02}$	$\mathbf{0.15_{\pm 0.02}}$

Table 5: Topographic similarity of emergent languages in the image discrimination game where images are compared with CelebA attributes. *No LN.* refers to the removal of the layer normalization on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to 0.01 ; *No LN. + WD* refers to the removal of layer normalization and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.

D.3 More Results with Listener Regularization

To complete the generalization scores of Table 1 in the main paper, we report in Table 6 the generalization scores in the image discrimination game for various regularization methods applied on the listener. We observe the same trends as in the reconstruction game. Indeed, listener regularization consistently improves the performances. It means, that a large gain of performance can be obtained in those games by regularizing the listener. The *Early stopping listener* remains a top line in image based experiments.

	Generalization ↑			Generalization ↑	
CelebA	$1/20$	$1/100$	**ImageNet**	$1/20$	$1/100$
Continuous	$0.67_{\pm 0.02}$	$0.39_{\pm 0.07}$	Continuous	$0.77_{\pm 0.01}$	$0.51_{\pm 0.03}$
No LN.	$0.67_{\pm 0.03}$	$0.44_{\pm 0.02}$	No LN.	$0.77_{\pm 0.01}$	$0.53_{\pm 0.03}$
No LN. + WD	–	$0.50_{\pm 0.07}$	No LN. + WD	$0.75_{\pm 0.01}$	$0.59_{\pm 0.04}$
Weight decay	$0.77_{\pm 0.04}$	$0.60_{\pm 0.06}$	Weight decay	$0.79_{\pm 0.03}$	$0.62_{\pm 0.02}$
Early stopping	$\mathbf{0.80_{\pm 0.03}}$	$\mathbf{0.69_{\pm 0.04}}$	Early stopping	$\mathbf{0.81_{\pm 0.01}}$	$\mathbf{0.64_{\pm 0.01}}$

Table 6: Comparison of generalization performances between the *Continuous listener*, *Early stopping listener* and listeners with regularization on the image discrimination game. *No LN.* refers to the removal of the layer normalization on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to 0.01 ; *No LN. + WD* refers to the removal of layer normalization and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.