Cultural Evolution Drives the Emergence of Conventions in Neural Emergent Communication

Maxime Toquebiau[1], Jérôme Botoko Ekila[1], Paul Van Eecke[1*], Katrien Beuls[2*]

[1]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Belgium
[2] Faculté d'Informatique, Université de Namur, Belgium
Correspondence: maxime.toquebiau@vub.be

Abstract

Recent studies of language evolution with neural agents often use models where agents can either speak or listen, where there are only two agents in the population, or where agents can share internal representations. These choices prevent the study of how language emerges through cultural evolution, i.e. through the cumulative adaptation of individual behaviours to fit the communicative needs of the community. We tackle this issue with a new model that trains populations of bi-directional neural agents in a fully distributed fashion to solve a referential game. We demonstrate the effectiveness of this approach on multiple image datasets and show that it is able to scale to large populations. We also show that more conventional languages are easier to learn, providing a direct advantage for cultural transmission. Overall, this work provides a novel approach to neural emergent communication that opens up new directions for evolutionary linguistics.

1 Introduction

Computational models of language evolution have played an insightful role in our understanding of the origins of human language. They allow us to test hypotheses on the roles of different environmental, cognitive or cultural biases in the emergence of natural language (Cangelosi and Parisi, 2002b; Kirby, 2002; Oudeyer and Kaplan, 2007; Steels, 2012). This approach is rooted in the view of languages as complex adaptive systems subject to evolutionary dynamics (Gell-Mann, 1992; Smith et al., 2003). Languages emerge and evolve to fit the communicative needs of populations of individuals (Dessalles, 2007). This can be modelled with artificial agents playing many interactions of language games (Steels and Vogt, 1997), progressively bootstrapping shared communicative strategies grounded in their observations.

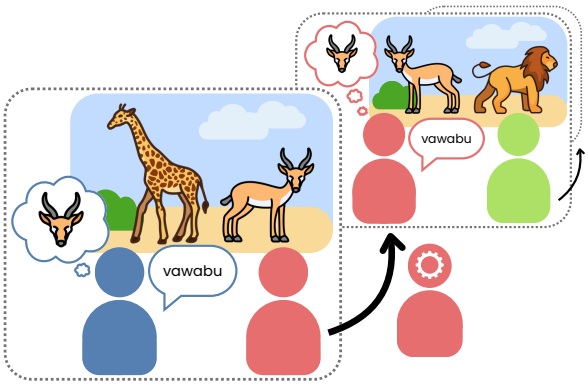

Figure 1: Cultural evolution of linguistic conventions for communicating about the world. Through repeated interactions with members of their community, individuals observe communicative behaviours, internalise them, and reuse them in later exchanges.

Over the past decade, the field of evolutionary linguistics has seen a new wave of interest with agents modelled as neural networks trained with deep learning techniques (Kottur et al., 2017; Lazaridou et al., 2017; Kharitonov et al., 2019; Chaabouni et al., 2020). Neural models of emergent communication allow us to scale language game experiments to more complex environments and languages (Lazaridou et al., 2018; Chaabouni et al., 2022). Previous work has used neural models to investigate specific mechanisms that may influence language emergence, including cognitive biases (Lian et al., 2021; Tucker et al., 2022), contextual information (Gualdoni et al., 2024; Kobrock et al., 2024) and population structure (Graesser et al., 2019; Rita et al., 2022a; Michel et al., 2023).

We are specifically interested in the mechanisms of cultural evolution that allow the emergence of natural language traits. One especially important trait is conventionality (Wright et al., 2026). Because linguistic constructions are arbitrary, they must rely on conventions to be of any use (Clark, 1996). These conventions emerge from the need to communicate about the world with various inter-

*Joint last authors.

Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing, pages 32202–32223
October 24-29, 2026 ©2026 Association for Computational Linguistics

locutors. As individuals take part in communicative interactions, communicative strategies are tested and selected for their ability to convey meaning to other individuals in the community (Steels, 2012). Individuals with cultural learning capabilities are able to observe and reproduce linguistic behaviours by understanding and internalising the intentional thinking behind their use (see Figure 1; Tomasello, 1999). This cultural evolutionary process is thought to be the main driver for the emergence of language in human populations (Tomasello, 2008). Thus, the question we tackle in this paper is the following: *Can we model the cultural evolution of linguistic conventions with neural agents?*

To this end, we propose a new method that enables a population of bi-directional agents to converge on a shared symbolic system for solving a reference game. Agents are trained to optimise mutual understanding in a fully decentralised way. We show that this model scales well to high-dimensional image datasets and large population sizes, and allows to study specific dynamics of cultural evolution. We show that the level of conventionality in a population is directly linked to how easy the language is to learn, thus favouring cultural transmission. Additionally, we demonstrate the value of learning from every individual interaction, instead of using batched interactions as is commonly done in the literature. Overall, this work provides a new way to model language emergence with neural agents, complementary to previous approaches, and should pave the way to more diverse experiments on neural emergent communication.

2 Related Work

Computational models of language emergence. Since the 1990s, researchers in computational linguistics, cognitive science and artificial intelligence have employed agent-based models to simulate the emergence of languages (Steels, 1995, 1997; Vogt, 2005). This was in reaction to the unavailability of historical records of such language emergence in humans – for obvious reasons. The synthetic ethology approach allows us to model the origins of linguistic traits by observing the emerging communicative strategies in populations of artificial agents (Cangelosi and Parisi, 2002a). This enables the implementation and testing of theories on the emergence of particular natural language traits such as compositionality (Brighton, 2002), word order (Steels and Garcia-Casademont, 2015; Van Eecke, 2018), case marking (Beuls and Steels, 2013; van Trijp, 2016), or speech (de Boer, 2000).

Central to this line of work is the idea that cultural evolution shapes the emergent languages. Because each individual in the population has its own knowledge of their previous communicative interactions, each new interaction becomes a transmission experience between two or more individuals. This has been extensively described in the iterated learning model (Smith et al., 2003; Smith and Kirby, 2008), where this cultural transmission process is often portrayed with successive generations of agents learning a language taught by previous generations from limited examples (Scott-Phillips and Kirby, 2010). This creates a transmission bottleneck that favours linguistic structures that are more learnable (Brighton et al., 2005). We aim to demonstrate this cultural transmission effect in populations of neural agents.

Cultural evolution in neural emergent communication. A large variety of neural models of language emergence have been proposed in the last decade. However, these models employ techniques that preclude capturing the dynamics of cultural evolution at play in language emergence (Galke et al., 2022): using agents that can only speak or listen (Chaabouni et al., 2022; Rita et al., 2022a; Michel et al., 2023; Lee, 2024), having populations of only two agents (Cao et al., 2018; Choi et al., 2018; Cheng et al., 2023; Rita et al., 2022b), sharing gradients or internal representations between agents (Tucker et al., 2022; Gualdoni et al., 2024), or having a combination of these design choices. These choices prevent fully capturing dynamics of cultural evolution, because cultural evolution specifically works in populations of distributed individuals able to try, observe and reproduce communicative behaviours (Tomasello, 1999). Notable exceptions are Graesser et al. (2019) and Lian et al. (2024) who both propose similar models to ours but use supervised learning on human language. Neural iterated learning models have also been proposed (Chaabouni et al., 2019; Ren et al., 2020), but focus on vertical transmission between successive one-agent generations.

Our work is the first to train populations of bi-directional agents in a fully distributed way and without human language supervision. This approach provides a new perspective on how cultural evolution drives the emergence of linguistic conventions in neural emergent communication.

3 Problem Definition

To assess how cultural evolution can lead to the emergence of linguistic conventions, we define our problem to explicitly consider the crucial elements that enable a cultural evolution process. Namely:

Populations. Cultural evolution occurs in communities of individuals interacting towards common goals (Mesoudi, 2011). Each individual participates by trying out strategies and observing the strategies of others. This leads to the evolutionary process in which, at the population level, unsuccessful strategies are discarded in favour of successful ones. Different population structures may affect the dynamics of this cumulative co-adaptation process (Raviv et al., 2019).

Situated local interactions. Communicative interactions take place locally between agents that observe the same scene – albeit from their own perspective (Beuls and Van Eecke, 2024). An interaction is characterised by the communicative need of agents to solve a particular task in their environment. By trying to solve this task, the agents test a strategy and evaluate it. Throughout many local interactions, agents accumulate experience and collectively converge towards conventions on how to communicate about entities of the world.

Distributed cognition. To evaluate realistic accounts of cultural evolution, cognition must be fully distributed. Agents each have their own internal representations and they learn only from their observations – sensory inputs, exchanged messages and feedback. Agents cannot access the internal representations of other agents, and there is no central controller orchestrating how agents should behave or what conventions should look like. The only shared signals are the symbols used for communication and the feedback to the success of interactions (Van Eecke et al., 2022). While this is a hard constraint for learning in a multi-agent system, this realistically models the strong pressure on symbolic systems to carry complex meaning through discrete symbols.

Bi-directional communication capabilities. Language evolution takes place through the successive sharing of cultural artefacts (Tomasello, 1999; Mufwene, 2019). This requires individuals to be able to both produce these artifacts and understand them (Pickering and Garrod, 2013). This interchangeability is a key property of human communication (Hockett, 1960) and languages have been shown to evolve through this ability to observe and reproduce communicative strategies (Andersen, 1973).

In this paper, our goal is to implement these elements of cultural evolution in a neural model of language emergence, and demonstrate how they provide sufficient pressure for the emergence of efficient linguistic conventions.

4 In-Context Reference Game

Our problem definition is instantiated in an *in-context reference game* where two agents are put in a shared scene and one agent – the speaker – needs to refer to an object in the scene by uttering a word to the other agent – the listener –, which then needs to point to the referred object. This setup has been used previously to highlight the role of contextual information in communication strategies (Steels and Kaplan, 1999; Botoko Ekila et al., 2024; Gualdoni et al., 2024; Główka et al., 2024; Kobrock et al., 2024; Zhang et al., 2025). The meaning of linguistic utterances depends on the context in which they are uttered (Wittgenstein, 1953; Grice, 1975). The in-context reference game allows us to model this important part of linguistic interactions.

Concretely, we define a population of N agents and a set of entities $W = \{e_1, ..., e_K\}$. An entity e_i is an object represented by a feature vector of dimension O. The game proceeds as a sequence of interactions. An interaction starts by sampling two agents from the population and assigning them the roles of speaker and listener, respectively denoted as S and L. The agents observe a scene $e = \{e_1, ..., e_C\} \subset W$, with C being the context size, i.e. the number of entities in the scene, and $e_i \in \mathbb{R}^O$. The topic of the interaction is always the last entity in the scene, i.e. $e_t = e_C$. The speaker observes the scene and produces a symbol w_S to discriminate the topic e_t from the other entities in the scene. The listener observes the scene, with entities randomly shuffled, and the uttered symbol w_S. Its goal is to infer which entity is the topic by adequately interpreting the speaker's message. If the listener's prediction is correct, agents are rewarded for a successful interaction. Otherwise, the interaction fails and the topic is disclosed to the listener. Based on this feedback, agents can update their strategies for future interactions.

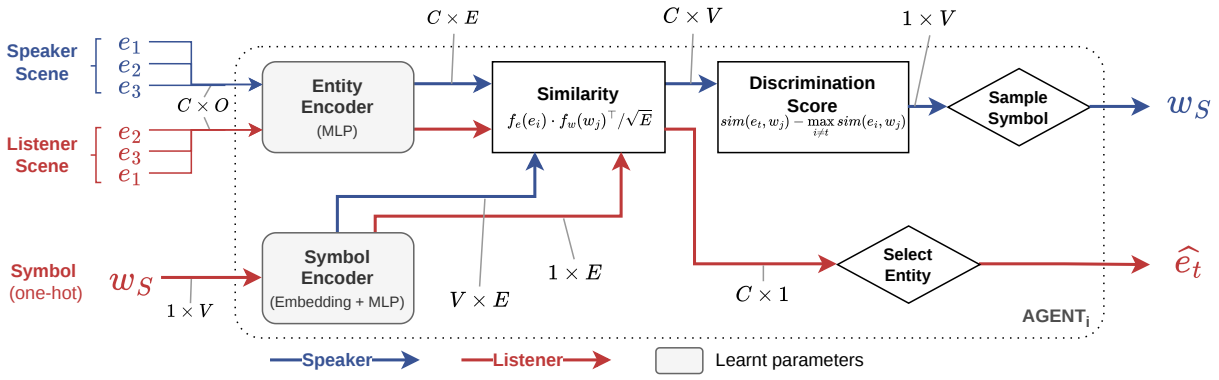

Figure 2: Architecture of an agent. The agent's Speaker and Listener forward passes are both represented by blue and red inputs and arrows. The learnt parameters are located in the two Encoders and are all used in both the Speaker and Listener roles. Dimensions of the matrices represented by arrows are shown.

5 Method

The goal of our methodology is to implement the necessary features for individuals in a population to enable the emergence of linguistic conventions through cultural evolution. That is, first, agents should be able to learn from every interaction and reuse the acquired knowledge in subsequent interactions, allowing a ratchet effect (Tennie et al., 2009) where agents progressively bootstrap the linguistic convention. This is achieved through a neural network architecture designed to have a small number of parameters that are updated after each interaction (see Section 5.1). Second, agents should have cultural learning capabilities allowing them to maximise mutual understanding (Tomasello, 1999, 2008). This is implemented as an imitation objective, similar to previous work (Cheng et al., 2023), where the listener agents explicitly learn to reproduce the intentional pattern observed during interactions (see Section 5.2).

5.1 Agent Architecture

Agents are modelled as neural networks able to learn *both* to produce symbols *and* understand symbols coming from other agents. This bi-directionality is exploited in the architecture to learn representations that serve both the production and interpretation tasks. Figure 2 illustrates the architecture of an agent and how it is used for speaking and listening. Agents are equipped with two distinct neural networks: an **Entity Encoder** f_e, consisting of a two-layer multi-layer perceptron (MLP), to encode observed entities; and a **Symbol Encoder** f_w, encoding symbols with an embedding matrix followed by a two-layer MLP (see Appendix A.1 for more details). These two encoders represent all the learnt parameters of an agent, denoted θ_i for agent i. Their output embeddings are used to produce and comprehend symbols, as described below.

Speaker role. As a speaker, the goal is to utter a symbol that allows the listener to understand which entity is the topic in the current scene. To do so, it must assess how well fitted each symbol in its vocabulary is to discriminate the topic in the scene. It receives as input the scene $e \in \mathbb{R}^{C \times O}$ with the topic e_t in the last position. Each entity in the scene e_i is encoded *separately* with the Entity Encoder, giving the entity embeddings $f_e(e|\theta_S) \in \mathbb{R}^{C \times E}$. Each symbol in the vocabulary is also encoded with the Symbol Encoder, giving the symbol embeddings $f_w(w|\theta_S) \in \mathbb{R}^{V \times E}$. From these two matrices, we compute the **similarity** between each entity and each symbol with a scaled dot-product:

$$\text{sim}(e, w|\theta_S) = \frac{f_e(e|\theta_S) \cdot f_w(w|\theta_S)^\top}{\sqrt{E}}, \quad (1)$$

with E the dimension of the embeddings. Scaling by $1/\sqrt{E}$ stabilises the magnitudes of the similarity values (Vaswani et al., 2017).

Equation 1 gives the matrix $\text{sim}(e, w|\theta_S) \in \mathbb{R}^{C \times V}$ containing all the similarities between each pair of entity and symbol. As in Botoko Ekila et al. (2024), we compute the **discriminative power** of each symbol w_i for the topic e_t, as the difference between the similarity of (e_t, w_i) and the maximum similarity between w_i and the other entities in the scene, i.e.:

$$\delta(e_t, w_i|\theta_S) = \text{sim}(e_t, w_i|\theta_S) - \max_{j \neq t} \text{sim}(e_j, w_i|\theta_S). \quad (2)$$

This discrimination score measures how much each symbol is "more similar to the topic than to the other entities". To make the discrimination score fully differentiable with regard to all entities in the scene, we replace the max operation by the continuous approximation $LogSumExp$ (LSE):

$$\begin{aligned}\delta(e_t, w_i|\theta_S) &= \mathrm{sim}(e_t, w_i|\theta_S) \\ &\quad - LSE_{j\neq t}\big(\mathrm{sim}(e_j, w_i|\theta_S)\big) \qquad (3)\\ &= \mathrm{sim}(e_t, w_i|\theta_S) \\ &\quad - \log \sum_{j\neq t} e^{\mathrm{sim}(e_j, w_i|\theta_S)/\tau_{lse}}, \qquad (4)\end{aligned}$$

with τ_{lse} a temperature parameter that can be used to make $LogSumExp$ closer to the max function (in practice we keep it at 1). Using the $LogSumExp$ allows us to consider all the entities in the scene to compute the discrimination score and provides richer feedback during learning updates.

The speaker's policy is the distribution over words in the vocabulary derived from the discrimination scores of all the words in its vocabulary:

$$\pi(w_j|e, \theta_S) = softmax\big(\delta(e_t, w|\theta_S)/\tau_s\big)_j, \quad (5)$$

with τ_s a temperature parameter controlling the flatness of this distribution, which can be used to control the level of exploration in the speaker's policy. The speaker's output symbol w_S is sampled from this distribution and sent as a one-hot vector to the listener.

Listener role. As a listener, agent L has to interpret the received symbol and point to the correct entity in the scene. As for the speaker, it encodes the scene e with its Entity Encoder: $f_e(e|\theta_L) \in \mathbb{R}^{C\times E}$. The received symbol w_S goes through the Symbol Encoder to get $f_w(w_S|\theta_L) \in \mathbb{R}^{1\times E}$. From these embeddings, the similarity $\mathrm{sim}(w_S, e_i|\theta_L)$ between the received word and each entity in the scene is computed as defined for the speaker. Based on these similarity values, the listener computes a probability distribution over the entities:

$$p_L(e_i|e, w_S, \theta_L) = softmax\big(\mathrm{sim}(e, w_S|\theta_L)\big)_i. \quad (6)$$

The listener then selects the entity with the highest similarity to w_S as its prediction $\widehat{e_t}$. If $\widehat{e_t} = e_t$, the listener has pointed to the right entity and the interaction is successful. Otherwise, the interaction is a failure.

5.2 Training

After each interaction, agents are trained to improve their success in upcoming interactions. The two participating agents receive a reward $R = 1$ for success or $R = 0$ for failure. In case of failure, the topic entity is also disclosed to the listener, allowing to compute a supervised learning objective. Based on this feedback and the events of the interaction, the speaker and listener agents compute their learning objective separately and update their parameters accordingly.

Speaker objective. The speaker's goal is to choose the words that will maximise the success of its interactions. This is defined as the REINFORCE objective (Williams, 1992):

$$\mathcal{L}_{speak}(\theta_S) = -\log\big(\pi(w_S|e, e_t, \theta_S)\big)R. \quad (7)$$

Minimising this objective will increase the probabilities assigned by $\pi(.|\theta_S)$ to successful symbols. The full derivation of this objective is provided and analysed in Appendix F.

In failed interactions, $R = 0$, so there is no update for the speaker. This shows an intended bias of this approach towards the speaking agent's choice of symbols to utter. The choices of speaker agents are considered valid, while not necessarily perfect.

Listener objective. The listener's primary goal is to interpret the received symbol correctly and point to the right entity. This is optimised by minimising the cross-entropy loss:

$$\mathcal{L}_{pred}(\theta_L) = -\log\big(p_L(e_t|e, w_S, \theta_L)\big). \quad (8)$$

Minimising $\mathcal{L}_{pred}$ will increase the similarity of the listener's representations of e_t and w_S.

In addition, the listener also tries to imitate the speaker's communication strategy by minimising the cross-entropy loss on its speaking policy $\pi(.|\theta_L)$, with the target output being the speaker's symbol:

$$\mathcal{L}_{imit}(\theta_L) = -\log \pi(w_S|e, e_t, \theta_L). \quad (9)$$

Because the speaking policy $\pi(.|\theta_L)$ uses the discrimination score (see Equation 5), the imitation objective provides a rich learning signal for the listener to update its representations.

The complete objective for the listener is therefore:

$$\mathcal{L}_L(\theta_L) = \beta_p\mathcal{L}_{pred}(\theta_L) + \beta_i\mathcal{L}_{imit}(\theta_L), \quad (10)$$

with β_p and β_i being two coefficient hyperparameters controlling the weights of each objective.

6 Experimental Setup

Datasets. We instantiate the in-context reference game in four different domains: the *ColorReference* dataset (Monroe et al., 2017), with 47K scenes of three colors; the *ManyNames* dataset (Silberer et al., 2020), as processed by Gualdoni et al. (2024), with 25K naturalistic scenes composed of an image with one topic object and one distractor (see the details of data extraction in the original paper); *CelebA* (Liu et al., 2015), with 202K images of celebrity faces; and *MSCOCO* (Lin et al., 2014), with 159K images of everyday scenes. For *CelebA* and *MSCOCO*, images are first encoded by a pre-trained DINOv2 (Oquab et al., 2024) model and scenes are constructed by randomly sampling C images and setting one as the topic of the interaction. Details about pre-processing and splits are available in Appendix B.

Training details. Agents are updated after each interaction they take part in. Training is fully decentralised. Unless specified, interactions are not batched, meaning that agents use only one interaction to compute their objectives. Our model uses $\beta_p = 1$ and $\beta_i = 0.2$, which were found to give the best trade-off between success and conventionality. Implementation details and other hyperparameters are provided in Appendix A.2. Our code is available at: gitlab.ai.vub.ac.be/mtoquebiau1/cultural-evo.

Evaluation. During evaluation, we evaluate the population by randomly sampling 1000 scenes from the held out test sets. All agents in the population are evaluated as speakers on these scenes, with randomly-selected listeners. We measure the average **communicative success** to track the efficiency of solving the game. To assess whether the evaluated populations converge to shared communicative strategies, we measure the level of **conventionality** of the emerged languages. We define the *population coherence* metric that looks at which symbols were used by the agents and computes the probability of having two agents utter the same symbol when observing the same scene (see Appendix C for detailed computation). Note that other metrics for conventionality have been proposed in the past for multi-symbol utterances (Raviv et al., 2019; Rita et al., 2022a; Michel et al., 2023; Kim and Oh, 2021). Our population coherence measure is equivalent to them for single-symbol utterances. Finally, we track the **number of distinct symbols** that account for 95% of all utterances, denoted $V_{used}^{95\%}$, to evaluate the size of the effective vocabulary in the emerged languages.

Method	Success (%)	Conv. (%)	$V_{used}^{95\%}$
ColorReference ($C = 3$, $O = 3$)			
Ours	$99.13_{\pm 0.15}$	$\mathbf{93.18}_{\pm 0.59}$	$8.8_{\pm 0.42}$
No-Imitation	$\mathbf{99.54}_{\pm 0.05}$	$76.46_{\pm 1.72}$	$20.5_{\pm 1.51}$
Similarity	$85.46_{\pm 0.78}$	$52.4_{\pm 9.17}$	$21.9_{\pm 4.25}$
S-L	$97.36_{\pm 0.35}$	$5.84_{\pm 2.6}$	$30.4_{\pm 2.41}$
ManyNames ($C = 2$, $O = 512$)			
Ours	$\mathbf{99.61}_{\pm 0.13}$	$\mathbf{94.11}_{\pm 0.37}$	$4.1_{\pm 0.32}$
No-Imitation	$99.11_{\pm 0.15}$	$53.81_{\pm 4.42}$	$13.6_{\pm 1.07}$
Similarity	$82.17_{\pm 11.74}$	$15.77_{\pm 8.14}$	$19.5_{\pm 3.95}$
S-L	$97.91_{\pm 0.2}$	$1.55_{\pm 1.24}$	$19.9_{\pm 1.2}$
CelebA ($C = 10$, $O = 384$)			
Ours	$97.41_{\pm 0.66}$	$\mathbf{82.74}_{\pm 1.17}$	$35.9_{\pm 2.51}$
No-Imitation	$\mathbf{98.63}_{\pm 0.28}$	$68.42_{\pm 1.8}$	$76.57_{\pm 5.68}$
Similarity	$38.07_{\pm 25.09}$	$15.51_{\pm 13.06}$	$29.2_{\pm 17.61}$
S-L	$88.54_{\pm 1.46}$	$13.83_{\pm 2.7}$	$62.3_{\pm 4.57}$
MSCOCO ($C = 10$, $O = 384$)			
Ours	$96.53_{\pm 0.5}$	$\mathbf{81.91}_{\pm 1.59}$	$32.5_{\pm 1.35}$
No-Imitation	$\mathbf{97.91}_{\pm 0.47}$	$66.51_{\pm 1.64}$	$77.6_{\pm 1.35}$
Similarity	$9.73_{\pm 0.24}$	$1.3_{\pm 1.53}$	$9.7_{\pm 1.16}$
S-L	$82.55_{\pm 0.85}$	$10.05_{\pm 1.28}$	$57.5_{\pm 2.32}$

Table 1: Results of our model and baselines on the evaluation sets of each dataset, after 2M training interactions, with 10 agents (10 runs each, mean and standard deviation shown). Context size C and entity dimension O are shown between parentheses.

7 Experiments

7.1 Model performance

We start by evaluating how our model performs in the four datasets and how it compares against two ablations: **No-Imitation** which does not use the imitation loss in the listener's objective, and **Similarity** which uses the similarity scores (Eq. 1) instead of the discrimination scores to compute the speakers' policy. In the Similarity baseline, speaker agents select their utterance based only on how similar symbols in their vocabulary are to the topic, thus not considering the other entities in the scene.

Table 1 shows the results after 2M interactions for populations of 10 agents. Our approach reaches near-perfect success on all datasets. The No-Imitation baseline is also highly successful with the highest score in three out of four settings. However, the conventionality of the emerged languages

is significantly higher for our model than without imitation. This shows that an explicit imitation mechanism helps neural agents converge towards common ground, as it is thought to occur in humans (Tomasello, 1999). The Similarity baseline performs particularly poorly in settings with a large context size C, showing that the in-context reference game benefits from the use of the discrimination score to efficiently learn context-based representations. It still obtains acceptable success rates on *ColorReference* and *ManyNames*, where C is small, but with significantly less conventionality than in our model. These results show that the discrimination score helps agents converge on similar representations by allowing richer learning updates.

7.2 Bi-directional vs. Speaker-Listener

Next, we demonstrate the importance of having bi-directional agents by introducing a **Speaker-Listener** (S-L) baseline where half of the population always act as speakers and the other half as listeners[1]. As shown in Table 1, the S-L baseline obtains high success on all datasets, though significantly lower than our model, but with low levels of conventionality. This indicates that S-L populations are able to solve the in-context reference game without converging to a single shared language.

To further visualise this, we compare the relative usage frequency of symbols across agents in our model and the S-L baseline, as shown in Figure 3. Both variants are trained on *CelebA* with 10 agents until reaching 95% success, requiring approximately 500K interactions for our model and 20M for S-L (see Appendix D.3 for details on the construction of this figure). Despite achieving similar communicative success, conventionality is substantially higher in our model (74% vs. 16%). Agents in our model converge towards highly similar usage distributions, whereas S-L speakers rely on different subsets of symbols. This highlights how, in S-L populations, the listeners learn to adapt to the individual language of each speaker. On the other hand, with bi-directional agents, the constant transmission of communicative behaviours leads to a single well-conventionalised language.

[1]The S-L population being $N/2$ speakers and $N/2$ listeners ensures that S-L agents have the same number of training updates as in our model. In our model, agents have a $2/N$ chance of being selected for an interaction (either as speaker or listener), and in S-L, speakers (resp. listeners) have a $\frac{1}{N/2} = 2/N$ chance of being selected.

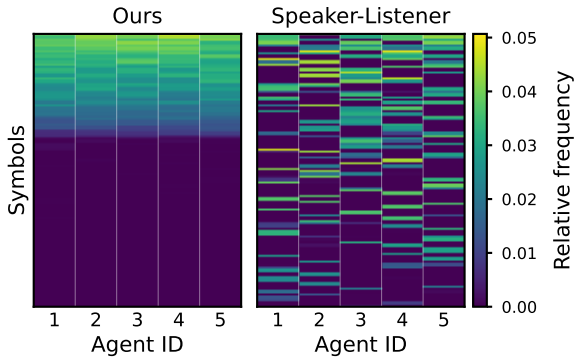

Figure 3: Relative usage frequency of symbols (rows) across 5 agents (columns) in the *CelebA* setup. Our model produces highly aligned symbol usage across agents, whereas the S-L baseline exhibits speaker-specific usage patterns, despite both populations reaching 95% success.

Note that the S-L baseline does not use the imitation objective as, conceptually, it does not make sense to train listeners to play the role of speaker. However, to ensure a fair comparison, we also provide results of the S-L baseline trained with the imitation objective on CelebA, with the same $\beta_i = 0.2$ as in our model, with 10 independent runs (see Appendix D.2 for further results with various values of β_i). Compared to our model, imitation has a similar but weaker effect on S-L. Success decreases slightly, from 88.54% to 86.14%. Conventionality is improved from 13.83% to 18.94%. This gain is smaller than the one observed from No-Imitation to our model (from 68.42% to 82.74%).

Overall, these results clearly indicate that S-L models do not produce significant pressure toward conventionality (see Appendix D.1 for results when varying V). S-L listeners are able to adapt to each speaker, even with a relatively small neural network architecture, and an imitation objective does not solve this issue. We show that having agents alternate speaking and listening roles, and learning representations that serve both of these roles, creates sufficient pressure toward conventionality. As listeners will act as speakers in future interactions, reproducing previously observed behaviours is now advantageous and successful behaviours will be selected, as in a typical cultural evolution process.

7.3 Scaling Populations

We investigate how well our model scales with the population size. Previous work has experimented with populations of up to 100 agents (Chaabouni et al., 2022). Thus, we compare populations of 10, 100 and 1000 agents on *ManyNames*. For a fair

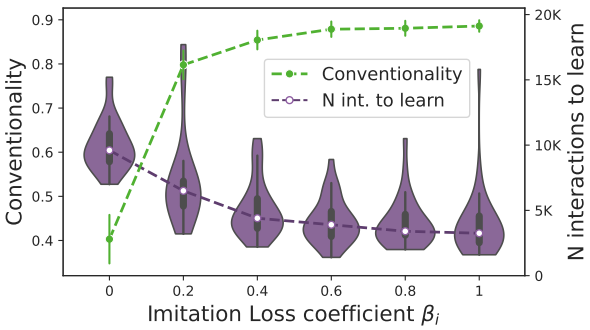

Figure 4: Learnability experiment in *ColorReference* with 100 agents, for different values of β_i (30 independent runs for each β_i). Green line shows mean conventionality when populations reach 95% success, with standard deviation in error bars. Violin plots show the distribution of the time to learn the language, with the dashed purple line indicating the median.

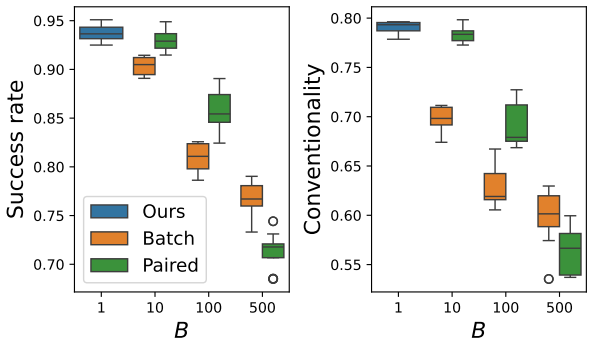

Figure 5: Success and conventionality after training on *CelebA* for 10M interactions, with 100 agents, comparing our model ($B = 1$) to "batched" and "paired" variants (10 runs each).

comparison, we train populations until they reach an average of 100K interactions per agent: 500K interactions for 10 agents, 5M for 100, and 50M for 1000 (see Appendix D.4 for more details and full results). Populations of 10, 100 and 1000 agents respectively obtain 98.97%, 97.4% and 95.89% average success and 91.23%, 91.62% and 92.01% average conventionality. The slight loss in success in larger populations can be understood as agents encounter more interlocutors and knowledge takes more time to be shared in the whole population. Nevertheless, this strong performance across all population sizes indicates that the model scales effectively.

7.4 Conventionality and Learnability

To highlight the role of conventionality in the cultural evolution of a language, we conduct a learnability experiment (Li and Bowling, 2019; Wright et al., 2026) to show that more conventional languages are easier to learn. We train populations of agents with different values of the imitation loss coefficient β_i, until they reach 95% success. Then, a new agent is added to the population and trained to understand the emerged language by playing as a listener (in all cases, the new agent is trained with $\beta_i = 1$). In Figure 4, we plot the conventionality of the emerged languages and the number of interactions for the new agents to reach 95% success, with 30 independent runs for each value of β_i.

These results first show that conventionality increases with β_i. We also observed a small loss in success when increasing β_i. This confirms a trade-off between conventionality and success, as already observed in the comparison with the No-Imitation baseline (see Table 1). Second, more conventional languages are easier to learn, going from 10K interactions to learn on average for $\beta_i = 0$ to less than 5K for $\beta_i = 1.0$. In conventionalised languages, the learner agent has fewer symbols to learn (see $V_{used}^{95\%}$ in Table 1) and observes them in action more consistently. This clearly highlights the reciprocal link between conventionality and transmission in cultural evolution, with each favouring the other.

7.5 Batched Interactions

Unlike previous work, our approach does not use batched interactions. Usually, agents interact over batches of scenes, with batch size B ranging from 32 to 1024, meaning that the two agents observe B different scenes, interact on all scenes independently, and finally compute an averaged optimisation objective given the experience gathered in the B interactions. Learning from a batch of samples is a common practice in deep learning to improve the stability of the learning updates. In our model, we do not use batched interactions to allow agents to learn from every interaction and immediately reuse the acquired knowledge in subsequent ones. To evaluate how this affects training, we compare our model ($B = 1$) against baselines using $B = \{10, 100, 500\}$. Since the learning rate may depend on batch size, we tuned it separately for each value of B, selecting the learning rate that achieves the best trade-off between success and conventionality (see Appendix D.5 for details).

Figure 5 shows the results on *CelebA*, with 100 agents, after training for 10M interactions. Training with batched interactions clearly leads to lower success and conventionality, with both decreasing as B increases. We hypothesise that the batched versions suffer from two problems: agents are up-

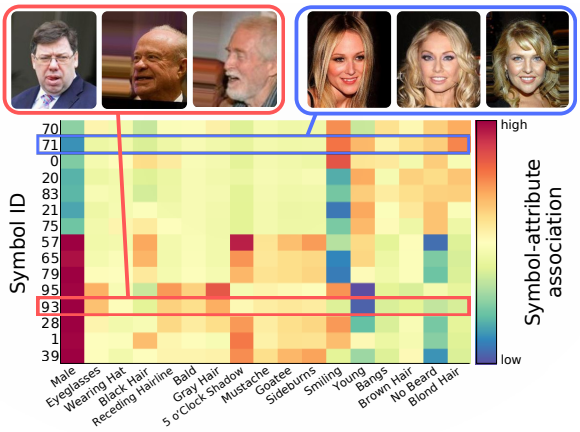

Figure 6: Co-occurrences between symbols and *CelebA* attributes for the 15 most frequently used symbols, computed from 100K interactions on unseen images (see Appendix D.6 for an extended version). Warmer colours indicate stronger positive associations between symbols and attributes. Symbols 71 and 93 are illustrated with images that most strongly elicited these symbols.

dated less often because interacting agents only update once after B interactions, and agents are paired for several interactions which may lead to more specialisation to one interlocutor. To investigate these issues, we experiment with an ablation where agents are also paired for $B > 1$ interactions, but are updated after each interaction with the same learning rate as our model. As shown in Figure 5, the success of "paired" baselines also decreases when increasing B. Interestingly, below $B = 500$, the paired baselines are better than the batched ones, showing that agents benefit from training after each interaction, even when they are paired. Overall, this highlights how changing partners and updating as often as possible are important to drive the cumulative bootstrapping of linguistic conventions.

7.6 Symbol Usage Analysis

To analyse the semantic structure of the emerged conventions, we evaluate trained agents on 100K interactions on unseen *CelebA* images and record the symbols produced. Although agents only observe the images, the dataset provides human-annotated visual attributes for each image (e.g. *wearing hat*, *smiling*), which we then use to plot the co-occurrences of these visual attributes and the symbols used by the agents (see Figure 6). We find that symbols capture bundles of correlated attributes, rather than isolated concepts. For example, symbol 71 is associated with *smiling*, *young* and *blond hair*, among others, whereas symbol 93 is associated with *male*, *gray hair* and not-*young*. These associations are also substantially more consistent across agents than those in the S-L baseline.

8 Discussion and Conclusion

We presented a neural approach to language emergence aimed at modelling the dynamics of cultural evolution through local situated interactions among agents in a population. This is achieved by modelling agents with bi-directional communication capabilities, where representations are shared for speaking and listening, and where agents are trained in a distributed fashion with an imitation objective inspired by cultural learning in humans. Under these conditions, agents are able to solve an in-context reference game in high-dimensional domains and across large populations.

The proposed model stands out for its sample efficiency, allowing agents to learn from every interaction and reuse acquired knowledge immediately after each interaction. This enables the ratcheting effect characteristic of cultural evolution, with agents adapting to each other and successful behaviours being selected at the population level. We showed that this does not happen in speaker-listener populations, where, instead, listeners adapt to the individual languages of every speaker agent. This clearly indicates that neural agents have the learning capacity to solve language games without cultural evolution. However, under the right conditions, cultural evolution does shape the emergent languages in populations of neural agents, by favouring conventionalised communicative behaviours. Ultimately, the emerged languages benefit from these conventions by having more compact, effective vocabularies and being easier to learn for new agents.

This work shows the importance of investigating the cultural pressures at play in language emergence, and provides methodological guidelines to successfully implement these pressures in neural models of emergent communication. In this respect, our approach supports recent efforts advocating for greater attention to the modelling assumptions and conditions under which languages emerge and evolve (Beuls and Van Eecke, 2024; Galke and Raviv, 2025). The proposed method is, therefore, complementary to previous work focused on cognitive biases in language use and calls for new work investigating the interplay between the cultural and cognitive pressures for language emergence.

Limitations

While this work explores the emergence of human-like languages through cultural evolution, it does not paint the full picture. One missing element is the ability to express meaning with multi-word utterances. Agents in our model only utter a single symbol. This is sufficient to study how conventions emerge for associating meaning from the world with symbolic representations. A natural extension of our model would modify the agent speaking policy with a recurrent neural network to allow the generation of sequences of symbols as utterances. This would allow us to model the emergence of compositional structures in utterances (Chaabouni et al., 2020; Galke et al., 2024).

For the sake of simplicity, we chose to focus on paired interactions. However, human communication can occur with more than two interlocutors, as explored previously with speaker-listener populations (Lee, 2024). Such grouped interactions may play an interesting role in the cultural evolution of language. We hypothesise that interactions involving one speaker and k listeners would improve the convergence speed and the conventionality level.

Finally, a limitation of our model is the fact that we need to set a fixed vocabulary size. This is common practice in neural emergent communication (Rita et al., 2025), because neural networks are usually initialised randomly at the start of each experiment. This introduces a bias that may have important effects on the results, as the size of the vocabulary determines the search space for the speaker policy. In this work, we chose to use $V = 100$ for all experiments, but other settings might require tuning V. This also limits our ability to study the role of invention of new symbolic forms to solve difficult communicative interactions (Tomasello, 1999; Skyrms, 2010). Here, invention occurs "by default" at the start by initialising random parameters for the Symbol Encoder of each agent. An interesting avenue would be to adapt the neural network architecture to allow new vocabulary entries to be added during the course of the game.

Acknowledgements

We are grateful to Liesbet De Vos, Lara Verheyen and Jamie D. Wright for their insightful feedback on the manuscript, and to the anonymous ARR reviewers, meta-reviewer and area chair for their constructive comments and suggestions. This research was supported by the F.R.S.-FNRS-FWO WEAVE project HERMES I under grant numbers T002724F (F.R.S.-FNRS) and G0AGU24N (FWO), the Flemish Government under the Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen programme and the AI Flagship project ARIAC by DigitalWallonia4.ai.

References

Henning Andersen. 1973. Abductive and deductive change. *Language*, 49(4):765–793.

Katrien Beuls and Luc Steels. 2013. Agent-based models of strategies for the emergence and evolution of grammatical agreement. *PLoS ONE*, 8(3):e58960.

Katrien Beuls and Paul Van Eecke. 2024. Humans learn language from situated communicative interactions. what about machines? *Computational Linguistics*, 50(4):1277–1311.

Jérôme Botoko Ekila, Jens Nevens, Lara Verheyen, Katrien Beuls, and Paul Van Eecke. 2024. Decentralised emergence of robust and adaptive linguistic conventions in populations of autonomous agents grounded in continuous worlds. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '24, pages 2168–2170, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Henry Brighton. 2002. Compositional syntax from cultural transmission. *Artificial Life*, 8(1):25–54.

Henry Brighton, Simon Kirby, and Kenny Smith. 2005. *Cultural selection for learnability: three principles underlying the view that language adapts to be learnable*, pages 291–309. Oxford University Press.

Angelo Cangelosi and Domenico Parisi. 2002a. *Computer Simulation: A New Scientific Approach to the Study of Language Evolution*, pages 3–28. Springer London.

Angelo Cangelosi and Domenico Parisi, editors. 2002b. *Simulating the Evolution of Language*. Springer London.

Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, and Stephen Clark. 2018. Emergent communication through negotiation. In *International Conference on Learning Representations*.

Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. 2020. Compositionality and generalization in emergent languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4427–4442. Association for Computational Linguistics.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. 2021. Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12):e2016569118.

Rahma Chaabouni, Eugene Kharitonov, Alessandro Lazaric, Emmanuel Dupoux, and Marco Baroni. 2019. Word-order biases in deep-agent emergent communication. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5166–5175. Association for Computational Linguistics.

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

Emily Cheng, Mathieu Rita, and Thierry Poibeau. 2023. On the correspondence between compositionality and imitation in emergent neural communication. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12432–12447. Association for Computational Linguistics.

Edward Choi, Angeliki Lazaridou, and Nando de Freitas. 2018. Compositional obverter communication learning from raw visual input. In *International Conference on Learning Representations*.

Herbert H. Clark. 1996. *Using Language*. ACLS Humanities E-Book. Cambridge University Press.

Bart de Boer. 2000. Self-organization in vowel systems. *Journal of Phonetics*, 28(4):441–465.

Jean-Louis Dessalles. 2007. *Why we talk: The evolutionary origins of language*. Oxford University Press.

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

Lukas Galke, Yoav Ram, and Limor Raviv. 2024. Deep neural networks and humans both benefit from compositional language structure. *Nature Communications*, 15(1).

Lukas Galke and Limor Raviv. 2025. Learning and communication pressures in neural networks: Lessons from emergent communication. *Language Development Research*, 5(1).

Murray Gell-Mann. 1992. *Complexity and Complex Adaptive Systems*, pages 3–18. Addison-Wesley, Redwood City, CA.

Laura Harding Graesser, Kyunghyun Cho, and Douwe Kiela. 2019. Emergent linguistic phenomena in multi-agent communication games. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, page 3700–3710. Association for Computational Linguistics.

H. P. Grice. 1975. *Logic and Conversation*, pages 41–58. Brill, Leiden, The Netherlands.

Eleonora Gualdoni, Mycal Tucker, Roger Levy, and Noga Zaslavsky. 2024. Bridging semantics and pragmatics in information-theoretic emergent communication. In *Advances in Neural Information Processing Systems*, volume 37, pages 21059–21078. Curran Associates, Inc.

Krzysztof Główka, Julian Zubek, and Joanna Rączaszek-Leonardi. 2024. Context-dependent communication under environmental constraints. *Cognitive Systems Research*, 88:101293.

Charles F. Hockett. 1960. The origin of speech. *Scientific American*, 203(3):88–97.

Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. 2019. EGG: a toolkit for research on emergence of language in games. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 55–60. Association for Computational Linguistics.

Jooyeon Kim and Alice Oh. 2021. Emergent communication under varying sizes and connectivities. In *Advances in Neural Information Processing Systems*, volume 34. Curran Associates Inc.

Simon Kirby. 2002. Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Kristina Kobrock, Xenia Isabel Ohmer, Elia Bruni, and Nicole Gotzner. 2024. Context shapes emergent communication about concepts at different levels of abstraction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3831–3848, Torino, Italia. ELRA and ICCL.

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

Dominik Künkele and Simon Dobnik. 2025. Learning to refer: How scene complexity affects emergent communication in neural agents. In *Proceedings of the 16th International Conference on Computational Semantics*, pages 289–297. Association for Computational Linguistics.

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

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

Heeyoung Lee. 2024. One-to-many communication and compositionality in emergent communication. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20794–20811. Association for Computational Linguistics.

Fushan Li and Michael Bowling. 2019. Ease-of-teaching and language structure from emergent communication. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Yuchen Lian, Arianna Bisazza, and Tessa Verhoef. 2021. The effect of efficient messaging and input variability on neural-agent iterated language learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10121–10129. Association for Computational Linguistics.

Yuchen Lian, Tessa Verhoef, and Arianna Bisazza. 2024. NeLLCom-X: A comprehensive neural-agent framework to simulate language learning and group communication. In *Proceedings of the 28th Conference on Computational Natural Language Learning*, pages 243–258. Association for Computational Linguistics.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. *Microsoft COCO: Common Objects in Context*, pages 740–755. Springer International Publishing.

Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 3730–3738.

Alex Mesoudi. 2011. *Cultural Evolution*. University of Chicago Press.

Paul Michel, Mathieu Rita, Kory Wallace Mathewson, Olivier Tieleman, and Angeliki Lazaridou. 2023. Revisiting populations in multi-agent communication. In *The Eleventh International Conference on Learning Representations*.

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

Salikoko S. Mufwene. 2019. *The Evolution of Language as Technology: The Cultural Dimension*, pages 365–394. University of Minnesota Press.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, and 7 others. 2024. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*.

Pierre-Yves Oudeyer and Frédéric Kaplan. 2007. Language evolution as a darwinian process: computational studies. *Cognitive Processing*, 8(1):21–35.

Martin J. Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347.

Limor Raviv, Antje Meyer, and Shiri Lev-Ari. 2019. Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, 286.

Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. 2020. Compositional languages emerge in a neural iterated learning model. In *International Conference on Learning Representations*.

Mathieu Rita, Paul Michel, Rahma Chaabouni, Olivier Pietquin, Emmanuel Dupoux, and Florian Strub. 2025. *Language evolution with deep learning*, pages 335–370. Oxford University Press.

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

Mathieu Rita, Corentin Tallec, Paul Michel, Jean-Bastien Grill, Olivier Pietquin, Emmanuel Dupoux, and Florian Strub. 2022b. Emergent communication: Generalization and overfitting in lewis games. In *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc.

Thomas C. Scott-Phillips and Simon Kirby. 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences*, 14(9):411–417.

Carina Silberer, Sina Zarrieß, Matthijs Westera, and Gemma Boleda. 2020. Humans meet models on object naming: A new dataset and analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1893–1905.

Brian Skyrms. 2010. *Signals: Evolution, Learning, and Information*. Oxford University Press.

Kenny Smith, Henry Brighton, and Simon Kirby. 2003. Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, 06(04):537–558.

Kenny Smith and Simon Kirby. 2008. Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509):3591–3603.

Luc Steels. 1995. A self-organizing spatial vocabulary. *Artificial Life*, 2(3):319–332.

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

Luc Steels. 2012. *Experiments in Cultural Language Evolution*. John Benjamins.

Luc Steels and Emilia Garcia-Casademont. 2015. How to play the syntax game. In *Proceedings of the European Conference on Artificial Life 2015*, pages 479–486. MIT Press.

Luc Steels and Frédéric Kaplan. 1999. Situated grounded word semantics. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 862–867. Morgan Kaufmann.

Luc Steels and Paul A. Vogt. 1997. Grounding adaptive language games in robotic agents. In *Advances in artificial life*, volume 4 of *Complex adaptive systems*, pages 474–482. MIT Press.

Claudio Tennie, Josep Call, and Michael Tomasello. 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528):2405–2415.

Michael Tomasello. 1999. *The Cultural Origins of Human Cognition*. Harvard University Press.

Michael Tomasello. 2008. *Origins of Human Communication*. Jean Nicod Lectures. MIT Press.

Mycal Tucker, Roger P. Levy, Julie Shah, and Noga Zaslavsky. 2022. Trading off utility, informativeness, and complexity in emergent communication. In *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc.

Paul Van Eecke. 2018. *Generalisation and Specialisation Operators for Computational Construction Grammar and their Application in Evolutionary Linguistics Research*. Ph.D. thesis, Vrije Universiteit Brussel.

Paul Van Eecke, Katrien Beuls, Jérôme Botoko Ekila, and Roxana Rădulescu. 2022. Language games meet multi-agent reinforcement learning: A case study for the naming game. *Journal of Language Evolution*, 7(2):213–223.

Remi van Trijp. 2016. *The evolution of case grammar*. Number 4 in Computational Models of Language Evolution. Language Science Press.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Paul Vogt. 2005. The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, 167(1–2):206–242.

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

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

Ludwig Wittgenstein. 1953. *Philosophical Investigations*. Wiley.

Jamie D. Wright, Lara Verheyen, Fabio De Ponte, Jérôme Botoko Ekila, Katrien Beuls, and Paul Van Eecke. 2026. Population-level linguistic conventionality benefits the robustness, learnability, and cognitive efficiency of emergent languages. *Journal of Language Evolution*, 11.

Yuqing Zhang, Ecesu Ürker, Tessa Verhoef, Gemma Boleda, and Arianna Bisazza. 2025. NeLLCom-Lex: A neural-agent framework to study the interplay between lexical systems and language use. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 10929–10945, Suzhou, China. Association for Computational Linguistics.

A Implementation details

A.1 Agent Architecture

The **Entity Encoder** and **Symbol Encoder** are the agent's neural networks. The Entity Encoder encodes each entity in a scene separately to an embedding with a multi-layer perceptron (MLP): $f_e : \mathbb{R}^O \rightarrow \mathbb{R}^E$. The Symbol Encoder encodes one-hot symbols with an embedding matrix followed by an MLP: $f_w : \mathbb{R}^V \rightarrow \mathbb{R}^E$.

The two MLPs in the encoders consist of two linear layers, each followed by a ReLU activation and a Layer Normalisation. They both use a hidden dimensionality H of 64. The embedding matrix in the Symbol Encoder maps one-hot vectors to embeddings of the same dimensionality as the hidden layers: $H = 64$.

This results in a rather lightweight architecture. For reference, a population of 10 agents playing on the *ManyNames* dataset ($O = 512$) with $V = 100$ comprises 522,240 learnable parameters.

Hyperparameter	Method			
	Ours	No-Imitation	Speaker-Listener	Similarity
Learning rate α			0.0001	
Vocabulary size V			100	
Hidden dimension H			64	
Embedding dimension E			64	
LogSumExp temperature τ_{lse}	1.0	1.0	1.0	-
Speaker policy temperature τ_s			0.1	
Prediction loss coefficient β_p			1.0	
Imitation loss coefficient β_i	0.2	0.0	0.0	0.0
Entropy max coefficient	-	-	-	0.1

Table 2: Hyperparameters used in all experiments, unless specified otherwise.

A.2 Training Details

Agents are trained after each interaction they take part in. To update the parameters, we use an RMSProp optimiser. The gradients of the speaker are always clipped so that their norm is no higher than 10. We found that it consistently led to better results. We hypothesise that this is because the update from REINFORCE is too noisy. All the hyperparameters we use are provided in Table 2.

As shown in the table, the Similarity baseline does not use the imitation objective. This is because we observed that imitation consistently led to the collapse of the Similarity agents' speaking policies to using only one symbol, thus leading to chance-level success. Unlike the main model and other baselines, it also uses an entropy maximisation objective (Williams and Peng, 1991) for the speaking policy to incite the exploration of different symbols. We found that it was required for the Similarity baseline to reliably converge to acceptable success rates. Note that the use of such an entropy maximisation objective is common in emergent communication (Chaabouni et al., 2022; Michel et al., 2023; Zhang et al., 2025). The fact that our model does not need it to solve the task indicates that the cultural evolutionary processes at play make policy exploration unnecessary.

Experiments were run on a CPU cluster equipped with *AMD EPYC 9535* processors. We estimate the total run time of our experiments to be around 480 hours.

B Datasets

The *ColorReference* dataset, consisting of human colour reference games over RGB colour patches, is used as provided by Monroe et al. (2017), without further pre-processing. For *ManyNames* (Silberer et al., 2020), a dataset of approximately 25K real-world images of everyday scenes, we follow the guidelines given by Gualdoni et al. (2024) to download the dataset and extract visual features with a pre-trained ResNet18 encoder (see their code), yielding feature vectors of dimension 512. All feature vectors are normalised between 0 and 1. For *CelebA* (Liu et al., 2015) and *MSCOCO* (Lin et al., 2014), we use the original train/test splits of the datasets. *CelebA* contains over 202K images of celebrity faces and *MSCOCO* (v2017) contains over 159K naturalistic images of everyday scenes. We extract visual features with a pre-trained DINOv2 (Oquab et al., 2024) model, yielding feature vectors of dimension 384. All feature vectors are normalised between 0 and 1. The pre-encoded and split datasets are available with the code online. All the datasets and models used are licensed under CC BY 4.0, except for DINOv2 which is licensed under Apache License 2.0, and CelebA which is "available for non-commercial research purposes only" as mentioned on their website.

C Conventionality Metric: Population Coherence

Conventionality in a given population can be measured by how coherent agents in the population would be when faced with the same scene and topic. Previously, this was estimated with the "linguistic coherence" or "pairwise coherence", in which agents in an interaction would be considered "coherent" if they would utter the same word to describe the topic.

This pairwise coherence can be extended to measure coherence at the population level. We define **population coherence** as the *probability of having two randomly-sampled agents utter the same word when given the same scene and topic*. This can be computed for any given scene $e = \{e_1, ..., e_C\}$, topic $e_t \in e$, and population P of N agents, by evaluating all agents in the population on the scene

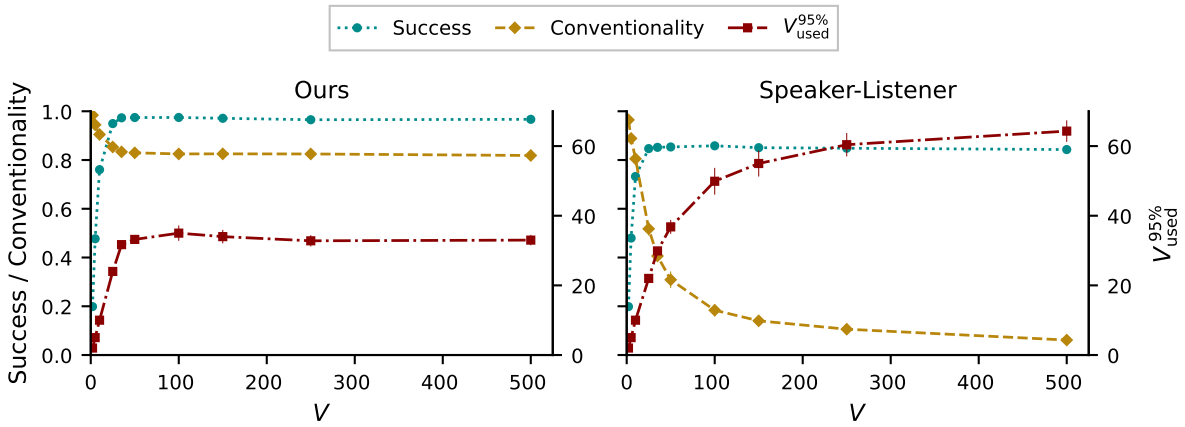

Figure 7: Effect of vocabulary size V. Small vocabularies ($V < 35$) under-capacitate both models, limiting communicative success. Once sufficient capacity is available, our model converges to a compact active vocabulary. For all $V \geq 35$, ~ 33 active symbols emerge, maintaining high conventionality. In contrast, the Speaker-Listener baseline gradually expands its effective vocabulary as V grows, while conventionality sharply decreases.

e. All words uttered w by the agents are stored and their usage u_w is computed as the number of times they were uttered by an agent. From this, we can compute population coherence with the following definition:

$$PC(e, e_t, P) = \sum_{w \in U} \big(p(\text{first agent utters } w) \times p(\text{second agent utters } w)\big),$$

with U the set of uttered words. The two probabilities are computed using the recorded usage of each word:

$$PC(e, e_t, P) = \sum_{w \in U} \frac{u_w}{N} \times \frac{u_w - 1}{N - 1} = \frac{1}{N(N-1)} \sum_{w \in U} u_w(u_w - 1).$$

This can be computed for a large set of scenes to measure the level of conventionality in the population.

D Details on Experiments & Additional Results

D.1 Varying the vocabulary size

Previous work has shown that constraints on the communication channel can shape emergent languages. In particular, discrete communication can favour lower-complexity systems (Chaabouni et al., 2021) and vocabulary size can affect communicative success and learnability (Künkele and Dobnik, 2025). This suggests a trade-off: restricting the available vocabulary may increase pressure towards a shared convention by forcing agents to reuse a smaller set of symbols, but an overly small vocabulary may not provide enough capacity to solve the task.

We therefore vary the vocabulary size V for both our model and the Speaker-Listener (S-L) baseline. In both settings, we train a population of 10 agents on *CelebA* with $V \in \{2, 5, 10, 35, 50, 100, 150, 250, 500\}$ for two million interactions. Figure 7 shows how the vocabulary size V affects communicative success, conventionality and the effective vocabulary size ($V_{\text{used}}^{95\%}$) in our model and the Speaker-Listener baseline.

In our bi-directional model, small vocabularies ($V \in \{2, 5, 10\}$) under-capacitate the agents, preventing them from solving the task. Up to $V = 10$, agents use the full available vocabulary, such that $V_{\text{used}}^{95\%} = V$. However, with $V = 2$, agents achieve only 19.84% success, and with $V = 10$, they achieve only moderate success (76.09%). At these vocabulary sizes, the languages are highly conventional ($> 90\%$), but the agents lack sufficient symbols to express the distinctions required by the task. Once the available vocabulary V is sufficiently large, the population consistently converges to a compact convention across the tested vocabulary sizes. For $V \geq 35$, agents maintain high success ($> 96\%$) and high conventionality ($> 81\%$), while consistently converging to an effective vocabulary of 31 to 35 symbols.

The Speaker-Listener baseline also suffers from under-capacity when V is too small ($V \in$

$\{2, 5, 10\}$), but it becomes increasingly sensitive to larger vocabulary sizes with respect to communicative success and conventionality. Larger vocabularies improve task success ($> 84\%$), but they sharply decrease population-level alignment. Conventionality drops from 80.51% at $V = 10$ to 18.46% at $V = 100$, and to only 6.14% at $V = 500$. At the same time, the effective vocabulary size ($V_{\text{used}}^{95\%}$) increases from 30 symbols at $V = 35$ to 64 symbols at $V = 500$. Thus, whereas our model stabilises around a compact effective vocabulary once sufficient capacity is available, the S-L baseline increasingly exploits the larger symbol space to the detriment of the emergent convention.

The results show that the S-L baseline is sensitive to a task-dependent parameter: the vocabulary size V must be large enough to solve the task, but small enough to maintain pressure towards conventionality. Crucially, the optimal vocabulary size cannot be determined a priori, as the number of required concepts to solve the task depends on the dataset and the number of entities in the scene. In contrast, our model consistently converges to a compact effective vocabulary once sufficient capacity is available.

D.2 Speaker-Listener with imitation

To fully compare our model to the S-L baseline, we provide results of S-L when trained with the imitation loss. Table 3 shows the results for three different values of the imitation loss coefficient $\beta_i \in \{0, 0.2, 1\}$, comparing the bi-directional and S-L models trained on CelebA for 2M interactions. In both cases, we observe a similar trade-off between success and conventionality controlled by the weight of imitation: increasing β_i improves conventionality but also hurts success. While conventionality can be improved by imitation in S-L, the gain is not substantial. This demonstrates that imitation is not sufficient to provide a pressure for conventionality and that bi-directionality is a crucial feature of our model.

D.3 Construction of Figure 3

We describe here the exact procedure used to construct Figure 3. The figure compares the relative usage frequency of symbols across agents in the *CelebA* setup with 10 agents and a vocabulary size of 100. We compare our model against the S-L baseline, where 5 agents acted as speakers and 5 as listeners. Both models were trained until reaching 95% communicative success, requiring approximately 500K interactions for our model and 20M for the S-L baseline. For each model, once training reached 95% success, training was stopped and the population was evaluated for an additional 100K interactions on the evaluation set of unseen images, during which we recorded the symbols that were produced. For readability, we consider the first five agents of our model (S-L having only five speaking agents), but the pattern of symbol usage is similar across the whole population. For every plotted agent, we count how often each symbol was produced in successful games. Counts are normalised independently for each agent such that every column represents the agent's relative distribution over produced symbols. Within each model, symbols are ordered by their total frequency across the plotted agents, with the most frequently used symbols appearing at the top.

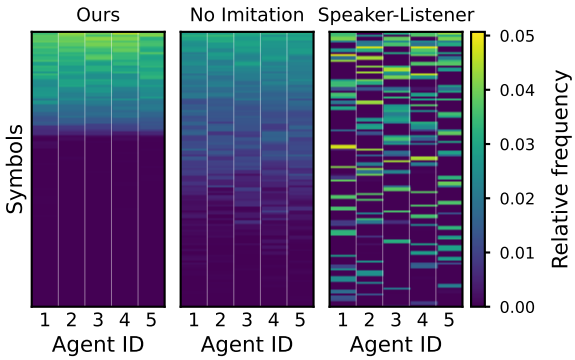

Figure 8: Relative usage frequency of symbols (rows) across 5 agents (columns) in the *CelebA* setup. Comparing our model (Ours) with the No-Imitation and Speaker-Listener baselines. All three variants are trained until reaching 95% success, but only our model produces highly aligned symbol usage across agents.

Finally, Figure 8 provides an extended version of Figure 3 including the No-Imitation baseline. The No-Imitation population was also trained until it reached 95% communicative success, which occurred after approximately 400K interactions. While the No-Imitation population exhibits partially overlapping symbol usage distributions, the degree of alignment remains substantially weaker than in our model.

D.4 Scaling Populations

Training larger populations inevitably requires conducting more interactions because each agent has a lower chance of being selected in every interaction, and thus of gaining experience. For a population of N agents, because each interaction selects a pair of

	β_i	Success	Conventionality
Bi-directional	0 (No-Imitation)	$98.63_{\pm 0.28}$	$68.42_{\pm 1.8}$
	0.2 (Ours)	$97.41_{\pm 0.66}$	$82.74_{\pm 1.17}$
	1	$89.25_{\pm 1.38}$	$87.3_{\pm 1.09}$
Speaker-Listener	0	$88.54_{\pm 1.46}$	$13.83_{\pm 2.7}$
	0.2	$86.14_{\pm 1.01}$	$18.94_{\pm 2.25}$
	1	$76.62_{\pm 2.22}$	$33.83_{\pm 5.8}$

Table 3: Results of the Bi-directional and Speaker-Listener models with different levels of imitation, controlled by β_i. Models were trained on CelebA for 2M interactions. Results are averaged over 10 independent runs and reported as mean $\pm$ standard deviation.

N	Total interactions	Success	Conv.
10	5×10^5	$98.97_{\pm 0.19}$	$91.23_{\pm 0.50}$
100	5×10^6	$97.40_{\pm 1.11}$	$91.62_{\pm 0.80}$
1000	5×10^7	$95.89_{\pm 0.16}$	$92.01_{\pm 0.28}$

Table 4: Results of scaling the population size after training for an equivalent of 100K interactions per agent on ManyNames ($C = 2$, $O = 512$). Results are averaged over 10 independent runs and reported as mean $\pm$ standard deviation.

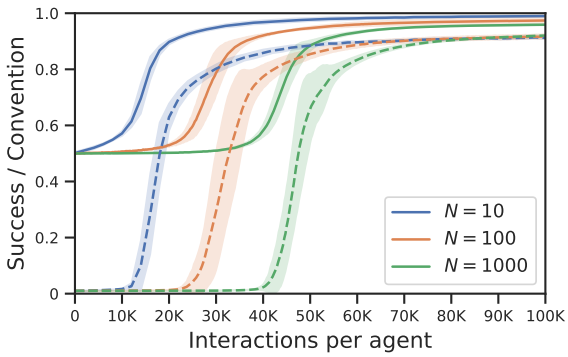

Figure 9: Training dynamics for populations of 10, 100 and 1000 agents on ManyNames. Curves show mean performance over 10 independent runs, with shaded regions indicating standard deviation. Solid lines indicate success, while dashed lines indicate conventionality.

agents, agents have a $2/N$ chance of being selected. Thus, we compare populations of different sizes with the same average number of interactions per agent.

Table 4 shows the number of interactions done for populations of 10, 100 and 1000 agents, and their performance on ManyNames after 100K interactions per agent. Figure 9 also shows the performance throughout training. It clearly shows the delayed convergence for larger populations. Because more agents participate in the bootstrapping of the emergent language, it takes more time for successful behaviours to be transmitted in the population and finally adopted by everyone. However, note that all populations seem to converge to the same levels of success and conventionality, showing that the model works well with large populations.

D.5 Batched Interactions

In the batch experiment of Section 7.5, we compare our model, which does not use batched interactions, with variants that use batch sizes of 10, 100 and 500. Increasing the batch size usually allows the learning rate to be increased, since the computed updates are more stable. Thus, before the experiment, we performed a parameter search for the learning rate. For each batch size B in $\{1 \text{ (ours)}, 10, 100, 500\}$ we tested the values $\{0.00008, 0.0001, 0.00015, 0.0002, 0.0003, 0.0005, 0.0008, 0.001, 0.0015, 0.002\}$. The learning rates that gave the best trade-off between success and conventionality were: for $B = 1$, $\alpha = 0.0001$; for $B = 10$, $\alpha = 0.0002$; for $B = 100$, $\alpha = 0.0005$; and for $B = 500$, $\alpha = 0.0008$. All "paired" baselines are trained with the same learning rate as ours, $\alpha = 0.0001$, because their agents are trained after each interaction.

D.6 Symbol Usage Analysis

Construction of Figure 6. In Section 7.6 of the main text, we introduced Figure 6. We describe here the exact procedure used to construct the symbol-attribute association heatmap and the accompanying example images. The figure was produced for our model in the *CelebA* setup with 10 agents and a vocabulary size of 100. The population was trained for 2M interactions. After training, the population was evaluated on an additional 100K interactions using the evaluation set of unseen images. During these interactions, we recorded the produced symbols together with the topic image and the distractor set shown in each game.

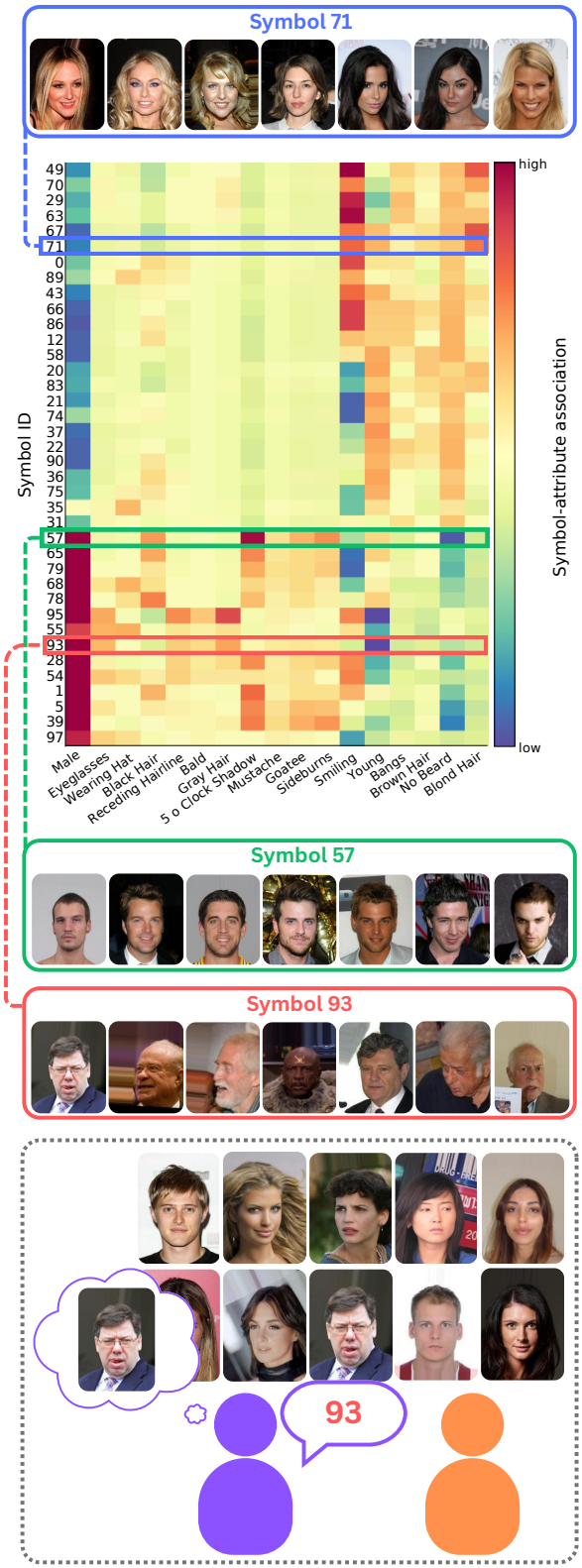

Figure 10: Extended version of Figure 6 from the main text showing co-occurrences between symbols and *CelebA* attributes for all 38 "active" symbols that were produced at least 50 times during 100K evaluation interactions on unseen images. Three symbols are highlighted with example topic images that elicited those symbols. At the bottom, an example interaction is shown in which symbol 93 was successfully used to discriminate the topic image from the distractor set.

The *CelebA* dataset contains 40 binary visual attributes annotated by humans for each image. These annotations describe visual properties such as whether the person is smiling, bald, blonde-haired, etc. Importantly, these attribute annotations are never observed by agents during training or evaluation and play no role in the learning process itself. They are used only afterwards as a post hoc analysis tool to interpret the semantic structure of the emerged symbols. Consequently, this analysis is restricted to the subset of properties explicitly annotated in the dataset metadata. The learned symbols may therefore also encode other latent visual regularities that are not captured by these annotations, such as background colour, head orientation or lighting conditions.

For every produced symbol s and every *CelebA* attribute a, we computed a contrastive association score measuring how much more frequently the topic image possessed the attribute compared to the distractor images in interactions where the symbol was produced. Positive values indicate that the symbol is produced more frequently when the topic image possesses the attribute than when the distractors do, whereas negative values indicate the opposite.

Due to space constraints and visualisation considerations, Figure 6 restricts the heatmap to the 15 most frequently produced symbols and the 17 comparatively salient and easily identifiable attributes, including facial hair, hair colour, accessories or age. Some of the excluded attributes are considerably more difficult to assess consistently from visual inspection alone, such as "*pointy nose*". Each row corresponds to a symbol ID and each column to a *CelebA* attribute. Each cell visualises the corresponding contrastive association score. The colour scale is centred around zero, with warmer colours corresponding to positive associations and colder colours to negative associations. Rows and columns were independently ordered using hierarchical clustering of the association scores using Euclidean distance. This clustering procedure was used solely for visualisation purposes to group similar rows and columns.

For symbols 71 and 93, we additionally display three example topic images sampled from evaluation interactions in which the corresponding symbol was produced by an agent. For each symbol, we ranked all corresponding interactions using the model's top-symbol discriminative score. The examples correspond to concrete cases in which the

symbol was successfully used to discriminate the topic from a particular set of distractors. We highlighted the three highest-ranking examples for each symbol.

Finally, Figure 10 provides a full version of the heatmap containing all 38 active symbols that were produced at least 50 times during the 100K evaluation interactions. In addition to the original examples, we highlight a third symbol (57), which is associated with younger males with light facial hair. For the three highlighted symbols, we display the seven highest-ranking examples rather than only the top three. These extended examples further illustrate that the learned symbols capture broad bundles of correlated properties. Finally, at the bottom of Figure 10, we provide an example interaction in which symbol 93 was successfully used. As shown, the symbol is used to discriminate the older gentleman (i.e. the topic of the interaction) from the relatively younger individuals in the distractor set.

E AI Assistant Use

All authors declare that this article was conceived and written without the use of generative AI tools, except to help write the code for generating the figures. They were *not* used to generate code used in the experiments.

F Derivation of the Training Objectives

F.1 Speaker objective

For a speaker, the objective is to maximise the success of the words it utters to describe entities. Here, we detail how the speaker, denoted S, computes its utterance distribution and how optimising this objective will impact the learnt representations.

The two networks to optimise are the entity encoder and the word encoder, with parameters $\theta_{S,e}$ and $\theta_{S,w}$, respectively. They produce the embeddings for observed entities and words in the vocabulary, $f(e_i|\theta_{S,e})$ and $f(w_j|\theta_{S,w})$, which will determine which word can be used for which entity. To this end, the basic component is the similarity score computed between these embeddings, defined as the scaled dot product:

$$\text{sim}(e, w|\theta_S) = \frac{f(e|\theta_{S,e}) \cdot f(w|\theta_{S,w})^\top}{\sqrt{E}}, \tag{11}$$

with $\theta_S = \{\theta_{S,e}, \theta_{S,w}\}$ and E the dimension of the embeddings. Given a scene $e = \{e_1, ..., e_C\}$ and the topic $e_t \in e$, the **discrimination score** of any word w_j can be computed from the similarity scores:

$$\delta(e_t, w_j|\theta_S) = \text{sim}(e_t, w_j|\theta_S) - \max_{i \neq t} \text{sim}(e_i, w_j|\theta_S). \tag{12}$$

This score measures the discriminative power of each word in the vocabulary w_j for the current scene and topic, by taking the similarity between w_j and the topic e_t, and subtracting the highest similarity between w_j and other entities in the scene. To allow this score to be fully differentiable with regard to all other entities in the scene, we can replace the $\max$ function by the differentiable estimate $LogSumExp$:

$$\delta(e_t, w_j|\theta_S) = \text{sim}(e_t, w_j|\theta_S) - LogSumExp_{i \neq t}\big(\text{sim}(e_i, w_j|\theta_S)\big) \tag{13}$$

$$= \text{sim}(e_t, w_j|\theta_S) - \log \sum_{i \neq t} e^{\text{sim}(e_i, w_j|\theta_S)/\tau_{lse}}, \tag{14}$$

with τ_{lse} a temperature parameter that can be used to make $LogSumExp$ closer to the $\max$ function (in practice we keep it at 1). The speaker's policy is the distribution over words in the vocabulary derived from the discrimination score:

$$\pi(w_j|e, e_t, \theta_S) = softmax\big(\delta(e_t, w|\theta_S)/\tau_s\big)_j \tag{15}$$

$$= \frac{e^{\delta(e_t, w_j|\theta_S)/\tau_s}}{\sum_k e^{\delta(e_t, w_k|\theta_S)/\tau_s}}, \tag{16}$$

with τ_s a temperature parameter controlling the flatness of this distribution, which can be used to control the level of exploration in the speaker's policy.

In each interaction, the speaker samples a word w_S from this distribution. The outcome of the interaction gives a reward R equal to either 1 for success or 0 for failure. From this reward, we compute the speaker's learning objective as the REINFORCE loss:

$$\mathcal{L}_{speak}(\theta_S) = -\log\big(\pi(w_S|e, e_t, \theta_S)\big)R. \tag{17}$$

Minimising this loss will increase the probability of selecting words that have led to success. To understand how this optimisation process impacts the word representations, we can compute the gradients of this loss with respect to the speaker's parameters to determine how the similarity scores are updated:

$$\nabla_{\theta_S}\mathcal{L}_{speak}(\theta_S) = -R\nabla_{\theta_S} \log \pi(w_S|e, e_t, \theta_S). \tag{18}$$

Let $z_j = \frac{\delta(e_t, w_j|\theta_S)}{\tau_s}$, for each word in the vocabulary w_j,

$$\nabla_{\theta_S}\mathcal{L}_{speak}(\theta_S) = -R\frac{\partial}{\partial z_j} \log \pi(w_S|e, e_t, \theta_S)\nabla_{\theta_S} z_j \tag{19}$$

$$= -R\big(\mathbb{1}\{j = S\} - \pi(w_j|e, e_t, \theta_S)\big)\nabla_{\theta_S} z_j \tag{20}$$

$$= R\big(\pi(w_j|e, e_t, \theta_S) - \mathbb{1}\{j = S\}\big)\nabla_{\theta_S} z_j. \tag{21}$$

Let $K = R\big(\pi(w_j|e, e_t, \theta_S) - \mathbb{1}\{j = S\}\big)$,

$$\nabla_{\theta_S}\mathcal{L}_{speak}(\theta_S) = K\nabla_{\theta_S} z_j \tag{22}$$

$$= \frac{K}{\tau_s}\nabla_{\theta_S}\delta(e_t, w_j|\theta_S) \tag{23}$$

$$= \frac{K}{\tau_s}\nabla_{\theta_S}\Big[\mathrm{sim}(e_t, w_j|\theta_S) - \log\sum_{i\neq t} e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}}\Big] \tag{24}$$

$$= \underbrace{\frac{K}{\tau_s}\nabla_{\theta_S}\mathrm{sim}(e_t, w_j|\theta_S)}_{D_t} - \underbrace{\frac{K}{\tau_s}\nabla_{\theta_S}\log\sum_{i\neq t} e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}}}_{D_{t^-}} \tag{25}$$

$$D_{t^-} = -\frac{K}{\tau_s}\nabla_{\theta_S}\log\sum_{i\neq t} e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}} \tag{26}$$

$$= -\frac{K}{\tau_s}\frac{1}{\sum_{i\neq t} e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}}}\sum_{i\neq t} e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}}\frac{1}{\tau_{lse}}\nabla_{\theta_S}\mathrm{sim}(e_i, w_j|\theta_S) \tag{27}$$

$$= -\frac{K}{\tau_s\tau_{lse}}\sum_{i\neq t}\left[\frac{e^{\mathrm{sim}(e_i,w_j|\theta_S)/\tau_{lse}}}{\sum_{k\neq t} e^{\mathrm{sim}(e_k,w_j|\theta_S)/\tau_{lse}}}\nabla_{\theta_S}\mathrm{sim}(e_i, w_j|\theta_S)\right] \tag{28}$$

$$= -\frac{K}{\tau_s\tau_{lse}}\sum_{i\neq t}\Big[softmax\big(\mathrm{sim}(e_{\neq t}, w_j|\theta_S)/\tau_{lse}\big)_i\nabla_{\theta_S}\mathrm{sim}(e_i, w_j|\theta_S)\Big] \tag{29}$$

The terms D_t and D_{t^-} describe how the similarities between observed entities in the scene and words in the speaker's vocabulary are updated. D_t concerns the similarities between the topic and all words, and D_{t^-} the similarities between all the distractors and all words.

Because the reward R is 0 in failure cases, $K = 0$ and there is no update in these cases. In other words, the speaker will not change its representations if the interaction has failed. This shows an (intended) bias towards the speaker's choices in this algorithm: when a speaker chooses a word to utter, this choice is considered valid (while not necessarily perfect) and the listener should try to understand it (and optimise for it).

In the successful case, $K = \pi_S(w_j) - \mathbb{1}\{j = S\}$ (for the purpose of explanation, we note $\pi(w_j|e, e_t, \theta_S)$ as $\pi_S(w_j)$), which gives the following updates:

- For $j = S$, i.e. the word uttered by the speaker, we have $K = \pi_S(w_S) - 1$. Because $\pi_S(w_S) \in [0, 1]$, then $K \leq 0$. Thus, minimising D_t will increase $\mathrm{sim}(e_t, w_S|\theta_S)$ proportionally to $\frac{|\pi_S(w_S)-1|}{\tau_s\tau_{lse}}$.
  For D_{t^-}, $-K = 1 - \pi_S(w_S) \geq 0$, thus minimising D_{t^-} will decrease each $\mathrm{sim}(e_i, w_S|\theta_S)$, with $i \neq t$, proportionally to $\frac{1-\pi_S(w_S)}{\tau_s\tau_{lse}} softmax\big(\mathrm{sim}(e, w_S)/\tau_{lse}\big)_i$.
  In other words, the update increases the similarity between the embeddings of the topic entity and the uttered word proportionally to the output probability of uttering this word (**reward successful word**), and decreases the similarities between the distractor entities and the uttered word proportionally to their uttering probability and to their relative weight in the similarity score (**punishing competitors**).

- For $j \neq S$, i.e. all the other words not chosen by the speaker, $K = \pi_S(w_j) \geq 0$. Minimising D_t will decrease the similarities $\mathrm{sim}(e_t, w_j|\theta_S)$, proportionally to $\frac{\pi_S(w_j)}{\tau_s\tau_{lse}}$.
  Minimising D_{t^-} with $-K = -\pi_S(w_j) \leq 0$ will increase the similarities $\mathrm{sim}(e_i, w_j|\theta_S)$, with $i \neq t$, proportionally to $\frac{\pi_S(w_j)}{\tau_s\tau_{lse}} softmax\big(\mathrm{sim}(e, w_j)/\tau_{lse}\big)_i$.
  In other words, this update decreases the similarity between the words that were not uttered and the topic entity proportionally to their uttering probability (**punishing competitors**), and increases the similarity between the words not uttered and the distractor entities proportionally to their uttering probability and to their relative weight in the similarity score (**realigning representations**).

F.2 Listener objective

As a listener, the objective is to correctly understand the speaker's utterance and to align its speaking policy with the speaker. In each interaction, the listener, denoted L, observes the scene $e = \{e_1, ..., e_C\}$ and the speaker's utterance w_S. It computes a distribution over the entities in the scene from which it samples its prediction (in practice, we always take the entity with the highest predicted probability). The distribution is computed directly from the similarity scores between the received word and each entity in the scene:

$$p_L(e_i|e, w_S, \theta_L) = softmax\big(\text{sim}(e, w_S|\theta_L)\big)_i. \tag{30}$$

The listener's first learning objective is to maximise the probability given to the topic, i.e. minimise the cross-entropy loss:

$$\mathcal{L}_{pred}(\theta_L) = -\log\big(p_L(e_t|e, w_S, \theta_L)\big). \tag{31}$$

With a computation similar to that for the speaker loss, this gives the following gradients:

$$\nabla_{\theta_L}\mathcal{L}_{pred}(\theta_L) = \big(p_L(e_i|e, w_S, \theta_L) - \mathbb{1}\{i = t\}\big)\nabla_{\theta_L}\text{sim}(e_i, w_S|\theta_L), \tag{32}$$

for each entity e_i in the scene.

A second learning objective is optimised by the listener to align its word usage with the speaker. This is done by updating the listener's speaking policy towards using the same word as the speaker to describe the topic in this scene. The probability of the speaker's word w_S in the listener's policy is:

$$\pi(w_S|e, e_t, \theta_L) = softmax\big(\delta(e_t, w|\theta_L)\big)_S. \tag{33}$$

The imitation objective maximises this probability, by minimising the following loss:

$$\mathcal{L}_{imit}(\theta_L) = -\log \pi(w_S|e, e_t, \theta_L). \tag{34}$$

This is almost the same objective as the speaker's $\mathcal{L}_{speak}$, without the reward R and with all temperature parameters set to 1. Thus, we can derive the gradients of $\mathcal{L}_{imit}$, for each word w_j, as:

$$\begin{aligned}
\nabla_{\theta_L}\mathcal{L}_{imit}(\theta_L) &= K\big(D_t - D_{t^-}\big),\\
\text{with } K &= \big(\pi(w_j|e, e_t, \theta_L) - \mathbb{1}\{j = S\}\big),\\
D_t &= \nabla_{\theta_L}\text{sim}(e_t, w_j|\theta_L),\\
\text{and } D_{t^-} &= \sum_{i\neq t}\Big[softmax\big(\text{sim}(e_{\neq t}, w_j|\theta_L)\big)_i \nabla_{\theta_L}\text{sim}(e_i, w_j|\theta_L)\Big].
\end{aligned} \tag{35}$$

From these equations, we can understand how minimising $\mathcal{L}_{pred}$ and $\mathcal{L}_{imit}$ aligns the listener's policy to the speaker's. Equation 32 shows that:

- For $i = t$, i.e. the topic, $p_L(e_t) - 1 \leq 0$, so minimising the objective will increase the similarity between the topic and the speaker's utterance.
- For $i \neq t$, i.e. all other entities in the scene, $p_L(e_i) \geq 0$, so the similarity between the distractor entities and the speaker's utterance will be decreased.

Equation 35 mirrors the update of the speaker:

- For $j = S$, i.e. the word uttered by the speaker, we have $K = \pi_L(w_S) - 1 \leq 0$, thus minimising D_t will increase the similarity between the topic and the speaker's word.
  For D_{t^-}, $-K = 1 - \pi_L(w_S) \geq 0$, thus minimising D_{t^-} will decrease the similarities between the distractor entities and the speaker's word.
- For $j \neq S$, i.e. all the other words not chosen by the speaker, $K = \pi_L(w_j) \geq 0$, thus minimising D_t will decrease the similarities between the topic and the words.
  Minimising D_{t^-} with $-K = -\pi_L(w_j) \leq 0$ will increase the similarities between the distractor entities and the other words in the vocabulary.