Emergence of Recursive Language through Bootstrapping and Iterated Learning

Vikas Kumar
TCS Research
Pune, India
kumar.vikas17@tcs.com

Ajin George Joseph
Indian Institute of Technology Tirupati
Tirupati, India
ajin@iittp.ac.in

ABSTRACT

Recursive structures are fundamental aspects of many human languages, allowing the embedding of concepts within other concepts. These structures are thought to be key factors in the expressiveness and flexibility of human communication. Such structures evolve through continuous and iterated learning, transmitted across generations via the bottleneck of language transmission. In this paper, we study language acquisition and the emergence of groundedness and recursive linguistic structures through neural iterated learning, where expressing a goal requires multiple levels of communication. We model this process as a language game within the framework of a decentralized, multi-agent deep reinforcement learning setting, where agents with local learning and neural cognitive faculties interact through a series of dialogues. Our examinations reveal the emergence of a shared depth-1 recursive language, where agents are able to acquire and generalize their bootstrapped language for expressing complex concepts.

KEYWORDS

Emergent communication; multi-agent reinforcement learning

ACM Reference Format:
Vikas Kumar and Ajin George Joseph. 2025. Emergence of Recursive Language through Bootstrapping and Iterated Learning. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 9 pages.

1 INTRODUCTION

The ability to articulate ideas using spoken language has stood as one of humanity's most intricate yet impactful communication tools. The humans developed and mastered this technique by forming a well-structured language that is well grounded in their environment. The current human language seems to be more refined than what it was centuries ago, which happened due to the continuous evolution of the language characterized by incremental mutations as successive generations acquire and adapt language. Embedding the intricate and ever-evolving nature of human language remains one of artificial intellegence's most significant challenges. While the linguistic structures that emerge among AI agents are not predefined, it is desirable for them to exhibit properties akin to those found in human language. Some notable emergent properties include syntactic structures [15], compositionality, word order, generalization, brevity, stability, statistical regularity, complexity, coherence, and linguistic divergence. The primary work done to study the computational simulation of evolution, origins, language acquisition, and other characteristics of human language is being done by [2, 11, 48, 49, 54, 56]. With the recent advancement in the field of deep learning [36] with respect to computational tractability, one could observe rigorous applications of deep learning and deep reinforcement learning in the context of language games [9, 27], especially referential/discrimination games [20, 29], reconstruction games [23], navigation/action games [22, 37] and visual communication games [41]. The initial array of work [14, 20, 28, 29, 51] focuses on developing a framework to study the establishment of a common grounding in an artificial environment where agents are motivated by achieving a shared goal. A few others study certain characteristics of natural languages, such as the symbolic grounding [26, 33, 37], compositionality [1, 26, 32, 37, 43, 55], generalization [3, 5], brevity, regularity [23, 45], the cultural and architectural transmission [10, 43], language structures through ease-of-teaching pressure [32] and networked communication [18]. Some of the recent works also provide deeper analysis pertaining to the nature and factors affecting the semiotic dynamics underlying the emergence of language and language constructs. [26, 44, 53] delve into the factors and constraints such as selectionist criteria, utility, informativeness, memory capacity, and learning capabilities that contribute to the development of compositionality and [13, 16, 17] analyze conditions, inductive biases [13] and intrinsic motivation required for the emergence of a coherent language. Another direction in which language emergence is being evaluated is along the dimension of scale [7, 46], where the correlation between language characteristics and system complexity and population dynamics is examined, while [30] incorporates pre-trained general language models to develop task-specific language models and [6] discusses the relation between the concepts complexity and message length.

In this paper, we wish to study whether the emergent language among a multiagent population can be grounded and recursive [19], a fundamental aspect of human language faculty [39]. For this purpose, we introduce a novel guessing game setting comprising of multiple agents with neural faculties, and they must utilize both sensory input, conceptualization, and verbal communication to establish a coherent communication language for achieving complex shared goals. The agents must learn to exploit the implicit recursive characteristics of the conceptualization to communicate the goals, which are designed in such a way that a final goal is determined through intermediate goals. The game involves two phases: (`i`) the emergence of a bootstrapped proto-language for communicating basic concepts; (`ii`) agents further engage in communicating complex

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

goals (containing intermediate goals), whose communication will require multiple folds of verbal expression combining intermediate primary messages recursively to identify the final target. Additionally, we study how such language evolves in a multi-generational setting, where agents learn and transmit language corresponding to nested concepts across generations through iterated learning.

2 PROBLEM FORMULATION

We model our language game as a partially-observable general-sum Markov game [34, 40, 47] since we aim for the emergence of symbolic structures through interactions among agents who possess the cognitive ability to extract and reinforce commonalities across multiple experiences. Agents are assumed to possess only partial observation of the environment, reflecting typical human experiences where individuals can only perceive their local surroundings and view the world in a simplified manner. The state of the environment at time step k is denoted by $s^{(k)} \in \mathcal{S}$, where $\mathcal{S}$ is the set of all the environment states. We let $o_i^{(k)} \in O$ be the partial observation of agent i, which is characterized by the function $f : \mathcal{S} \mapsto O$, where O is the set of all possible observations. At time instant k, agent i chooses a random action $a_i^{(k)}$ which is dependent on the current observation according to a parameterized stochastic policy $\pi_{\theta_i}(\cdot|o_i^{(k)})$ which is a conditional probability mass function over action space $\mathcal{A}$ conditioned on the observation $o_i^{(k)}$. For agent i, each state transition yields a random reward $r_i^{(k)}$ according to the function $\mathcal{R}\colon \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto \mathbb{R}$. The system evolution is stochastic in nature and characterized by the probability transition function $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \mapsto [0, 1]$, where $\mathcal{P}(s, a, s') = \mathbb{P}r(s^{(k+1)} = s'|s^{(k)} = s, a^{(k)} = a)$ which is the conditional probability of next state is s' conditioned on the current state and action being s and a respectively. The collective goal of the agent population is to collaboratively seek a policy $\pi_{\theta*} = [\pi_{\theta_1^\star}, \pi_{\theta_2^\star}, \ldots, \pi_{\theta_N^\star}]$ that maximizes the individual long-term return over the network based solely on local information, *i.e.*,

$$\theta_i^\star = \arg\max_{\theta\in\Theta} J_i(\theta), \text{ with } J_i(\theta) = \mathbb{E}_{\boldsymbol{\pi_\theta},\mu}\left[\sum_{k=0}^{T-1} \mathbf{r}_i^{(k)}\right]. \quad (1)$$

where $\mathbb{E}_{\boldsymbol{\pi_\theta},\mu}[\cdot]$ is the expectation with respect to all T length trajectories generated using the policy π_θ with initial distribution μ and $\Theta \subset \mathbb{R}^s$ is a compact and convex set.

3 LANGUAGE GAME

In this paper, we consider a "guessing game" environment [50] for studying emergent recursive language, which can be understood as the complexity addition to the widely used referential game (Lewis signaling game) [31]. In referential games, the listener and speaker can immediately reach an agreement on the meaning topic of an unknown utterance. However, in guessing games, multiple distinct concepts are present in the context during the interaction. These additional elements, along with the topic, introduce further ambiguity for the agents to find an agreement for an unknown word. This is also known as Quine's "Gavagai" problem [42] or poverty of stimulus. This complexity increases even further when the number of concepts under a category increases due to the combinatorial explosion of mappings from words to meaning.

In this paper, we introduce a **PathFinder** game that consists of a simple connected graph $G = (V, E)$ on a bounded $2D$ plane, where V and E are vertices and edges, respectively. Although similar in shape, each vertex is distinct by a unique location ($\in \mathbb{R}^2$) and a color is assigned to them from a predefined set of colors $\mathcal{B}$. The agents residing in the environment conceptualize the topology of the environment into concept space, $C = \mathcal{H} \cup \mathcal{W} \cup \mathcal{B} \cup \{\square\}$ which consists of a finite collection of segments $\mathcal{H}$, sectors $\mathcal{W}$, colors, $\mathcal{B}$ and the null concept $\square$. A segment is defined as the circular strip formed by concentric circles centered at a certain point. Sectors are formed by uniformly dividing the 2D region, *i.e.*, at 90° angles (quadrants) and 45° angles (octants). Please see Figure 1 for illustration. This conceptualization parallels how humans perceive locations using cardinal directions. In our setting, the space is conceptualized a priori as discrete and categorical. The color property assigned to each node is based on the assumption that agents possess a sensory mechanism to identify color hues. Hence, each node can be characterized as the $\langle \mathcal{H}, \mathcal{W}, \mathcal{B}\rangle$. The sectors belonging to $\mathcal{H}$ may overlap with each other, potentially leading to a poverty of stimulus scenario. The language game proceeds as follows:

(1) A speaker and listener are chosen from the population U and assigned to a random source vertex s. Speaker picks a random target vertex t from graph G, which is at least one unit hop away from the source. The speaker's goal is to guide the listener to the target vertex t from the source vertex s through a finite path $P = \langle v_0 = s, v_1, v_2, ..., v_h = t\rangle$ by uttering a finite sequence of messages.
(2) The speaker agent sequentially obtains the observation $o = \langle o_0, o_1, \ldots, o_h\rangle$ corresponding to the path P, where $\boldsymbol{o}_i = f(v_i) \in \mathbb{R}^d$. The concept selection module generates the concept $c = \langle c_0, c_1, \ldots, c_h\rangle$ based on the observation o by adhering to minimum description length criteria to identify the essential concepts for conveying the target vertex t. Here, the elements of the concept c are interrelated as $c_i = \Lambda(o_i, o_{i-1})$.
(3) The utterance module of the speaker sequentially produces the messages $m = \langle m_0, m_1, \ldots, m_h\rangle$ corresponding to the selected concepts, where $m_i \in \Xi$ (vocabulary).
(4) The listener hears the sequential message m uttered by the speaker and predicts the concepts $c' = \langle c'_0, c'_1, \ldots, c'_h\rangle$ corresponding to the message m and further deduces the path P' to the target vertex t'. Then the listener navigates to t' by following the predicted path P'. Note that the listener does not have access to the concept-utterance mapping of the speaker.
(5) The listener and speaker are awarded a shared reward if $t = t'$ the listener identifies the target vertex). Steps 1–5 constitute one conversation in a dialogue, which is repeated for all conversations within the dialogue.
(6) After a predetermined number of dialogues, the successive generation of language-innocent agents U' is added to the population, *i.e.*, $U = U \cup U'$, and then the game resumes.

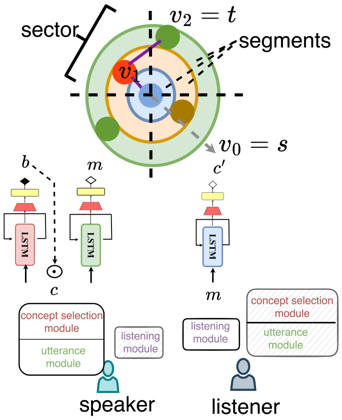

Figure 1: Illustration of the PathFinder game with homogeneous agents

3.1 Recursive Language Structure

The recursive structure refers to a syntactic or structural pattern in language where an element is repeated in a self-referential manner within a sentence or a larger linguistic unit. The recursion is one of the fundamental and innate properties of the language faculty, which plays a crucial role in the generation of complex and hierarchically structured expression [19]. "*A language that lacks recursion would be considerably . . . exotic. No sentence in such a language could contain more than two words*" [38]. In linguistics, recursive language enables the creation of sentences or phrases of arbitrary complexity through repetitive structures generated by a grammatical mechanism. In human language, situations are often expressed by embedding one concept within another. For example, in the sentence "*The dog the cat bit barked*", can be marked as a recursive sentence as follows: "[[*the dog* [*the cat bit*]] *barked*]". Recursion, as exemplified here, is a fundamental linguistic feature that allows "digital infinity", *i.e.*, the generation of infinitely complex expressions from a finite set of rules. It serves as a cornerstone of human language, facilitating the conveyance of intricate meanings and ideas. The recursive structures inherent in language greatly enrich and enhance the flexibility of expression in natural languages. Recursion, at *depth-n*, entails embedding a maximum of $n + 1$ components recursively within a single sentence structure. In this paper, our primary objective is to observe whether the emergent communication protocol can gravitate towards a recursive language while describing complex goals. The recursive language structure used by the agents in the PathFinder game can be represented using following recursive grammar:

$$
\begin{aligned}
\textit{Nonterminals}: \;& S, P \quad \textit{Start Symbol}: \; S \\
\textit{Rules}: \;& S \longrightarrow PS|\square \\
& P \longrightarrow C_1C_2C_3 \\
& C_1 \longrightarrow c_1, \forall c_1 \in \Gamma(\mathcal{H} \cup \{\square\}) \\
& C_2 \longrightarrow c_2, \forall c_2 \in \Gamma(\mathcal{W} \cup \{\square\}) \\
& C_3 \longrightarrow c_3, \forall c_3 \in \Gamma(\mathcal{B} \cup \{\square\})
\end{aligned}
$$

The language defined above is right recursive, where each phrase (also referred to as "derivational layer") forms the atom for the subsequent derivational layer. In this way, one can parse the recursive message in an iterative manner. However, within the phrase $P = C_1C_2C_3$, we have `sector` (C_1) followed by `segment` (C_2) and then `color` (C_3). Importantly, there is no recursion within the phrase itself; instead, it follows a hierarchical structure based on the order of sector, segment, and color. This structure makes the language compositional at the phrase level while allowing for recursive relationships between phrases. In this paper, our objective is to promote interactions between the agents, where they learn to communicate through decentralized learning by experimenting their knowledge of the nearly raw language through dialogues, which eventually results in the emergence of a shared vocabulary.

Definition (Emergent vocabulary): An emergent vocabulary Γ is a shared mapping function between lexis Ξ and the concept space C, *i.e.*, $\Gamma : \Xi \leftrightarrow C$ collectively agreed upon by all the agents in the population [12]. Note that there are $|C|^{|\Xi|}$ possible vocabularies for all the agents to agree upon, which makes it unlikely for all agents to converge on the same vocabulary without some mechanism for coordination and consensus.

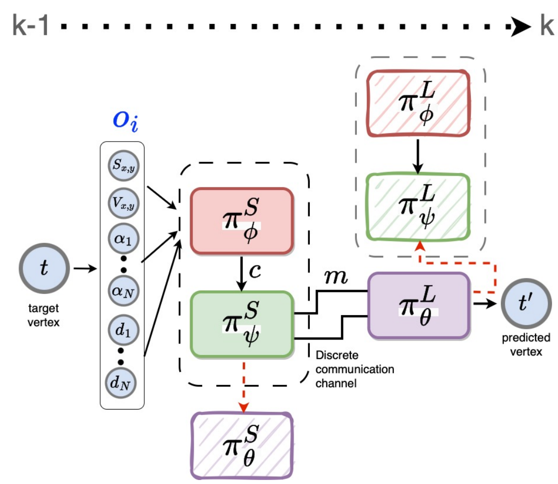

Figure 2: Transition of one-step communication between speaker and listener.

3.2 Agent Architecture

In our setting, each agent's architecture is identical and comprises of two modules: (`i`) the speaking module (concept selection module and utterance module) and (`ii`) the listening module. We call this combined policy architecture of all the modules residing in one agent a homogeneous agent. A homogeneous agent possesses the interchangeability property of human language. The concept selection module in the speaking module regulates the principle of least effort [4, 57] and allows the agent to optimize the decision of choosing the number of required spatial concepts for efficiently conveying the topic vertex. The utterance module is responsible for choosing a word according to the chosen concept by following a policy. Similarly, the listening module processes the uttered concepts by the speaker and interprets them to identify the conveyed goal. The utterance module and the concept selection module combine to form the speaking module, which involves decision-making and then uttering words. Each module is implemented using recurrent neural networks to handle sequential actions, such as choosing the relevant concepts, uttering a sequence of symbols, or processing a sequence of uttered symbols. The networks build internal representations that embody the tacit competence and understanding of the language. The combined architecture is termed as the policy architecture, where all the modules work in congruence to achieve a stable agent that has synchronization between all it's modules. The synchronization here means that the agent should learn the language, *i.e.*, mappings from concepts to utterance symbols and vice versa.

Here ϕ, ψ, and θ represent the parameters of the concept-selection network, utterance network, and listening network, respectively. All the modules of listener and speaker have to synchronize through trial and error for a successful communication language to emerge. In our setting, we perform decentralized learning with decentralized execution [14]. Our agents are independent learners [52] and the channel between speaker and listener is non-differentiable, which implies that the back-propagation of the listener does not transmit the gradient backward to the speaker. In our 2D environment, there are N agents and M vertices. The state $\mathcal{S}$ of the game set consists of all relevant details that define the environment. The state of the environment at time k is given by $\mathbf{s}_k = \left[\mathbf{x}^{(1),\dots,(N)}, \mathbf{z}_k^{(1),\dots,(N)}, \mathbf{q}^{(1),\dots,(N)}, \mathbf{u}_k^{(1),\dots,(N)}\right]^\top \in \mathcal{S}$, where $\mathbf{x}^{(i)} \in \mathbb{R}^2$ is the location of the i^{th} vertex in the world, $\mathbf{z}^{(i)} \in \{1, 2, \dots, N\}$ is the current location of the agent, i, $\mathbf{q}^{(i)} \in \mathbb{R}$ is the color of the vertex, i and $\mathbf{u}_k^{(i)}$ is the utterance in the conversation involving agent i. The speaker agent i locally perceives the environment, which characterizes the observation vector of the speaker agent $\mathbf{o}_k^{(i)} \sim \left[\mathbf{z}_k^{(i)}, \mathbf{g}_k^{(i)}, \mathbf{u}_k^{(i)} \mathbf{q}^{(\mathbf{g}_k^{(i)})}, \mathbf{d}^{(1),\dots,(M)} + \epsilon_d, \quad \mathbf{w}^{(1),\dots,(M)} + \epsilon_w\right]^\top$, where $\epsilon_d \sim \mathcal{N}(0,1)$ and $\epsilon_w \sim \mathcal{N}(0,1)$ are white Gaussian noises, $\mathbf{g}_k^{(i)} \in \{1, 2 \dots M\}$ is the topic vertex, and $\mathbf{d}, \mathbf{w}$ represent the distance and the angle of vertices from the speaker's current vertex respectively. The interaction pathway consists of multiple networks across the speaker and listener agents operating sequentially. The concept-selection network π_ψ operates in a one-to-many mode, where the initial hidden vector is obtained through a linear transformation of the observation vector $\mathbf{o}_k$, and the output is fed back as input. This network outputs the conception-selection bit-vector, $\mathbf{b}_k$ which is then passed through a differentiable channel to the speaking module π_θ (many-many mode) along with the spatial description $\mathbf{c}_k$ of the topic vertex as $\mathbf{c}_k \odot \mathbf{b}_k$, where $\odot$ is the coordinate-wise vector product. The network utters the message, $\mathbf{m}_k$ which is transmitted to the listener through a non-differentiable, noise-free channel. The listening module π_ϕ in the listener agent operates in a many-to-many mode, which means it processes the words in the generated message $\mathbf{m}_k$ sequentially and generates a probability distribution $\pi_\phi(\cdot|\mathbf{m}_k)$ over the entire concept space C. This distribution represents the agent's interpretation of the message in terms of different concepts within the concept space. This distribution is further used to generate the listener interpretation $\mathbf{c}'_k$ through categorical sampling. The complete architecture of the agent is depicted in Figure 1.

4 POLICY OPTIMIZATION

The cost function $J(\theta, \psi, \phi)$ combines three different functionalities involved in the processes, namely the interchangeability function, the concept selection module, and regularized communication between agents and training, which requires the joint optimization of all the components. Since gradients cannot be backpropagated through the discrete channel between concept selection and utterance network, we use Gumbel-Softmax [21, 35] based sampling to enable differentiability between concept-selection and utterance channel, allowing gradients to flow through the samples.

Then $J(\theta, \phi, \psi) = \mathcal{L}_1(\theta, \phi, \psi) + \mathcal{L}_2(\theta, \phi, \psi) + \mathcal{L}_3(\theta, \phi, \psi)$, where $\mathcal{L}_1(\theta, \phi, \psi) =$

$$\underbrace{\mathbb{E}_{\mathcal{I}_k}\left[\sum_{k=0}^{K-1} \mathbf{r}_k + \beta \mathcal{H}(\pi_{\theta^A}(\cdot|\mathbf{c}_k)) + \beta \mathcal{H}(\pi_{\phi^B}(\cdot|\mathbf{o}_k))\right]}_{\text{Regularized cumulative reward}}, \beta \geq 0,$$

$$\underbrace{\mathcal{L}_2(\theta, \phi, \psi) = -\mathbb{E}_{\mathcal{I}}\left[\|\mathbf{b}\|_2^2 + \beta' \mathcal{H}(\pi_{\psi^A}(\cdot|\mathbf{s}))\right]}_{\text{concept-selection loss}}, \beta' \geq 0.$$

$$\underbrace{\begin{aligned}\mathcal{L}_3(\theta, \phi, \psi) = -\mathbb{E}_{\mathcal{I}}\Big[&\alpha_1 \mathcal{D}_{KL}\left(\pi_{\theta^A}(\cdot|\mathbf{m}) \,\Big\|\, \pi_{\phi^A}(\cdot|\mathbf{m})\right) + \\ &\alpha_2 \mathcal{D}_{KL}\left(\pi_{\phi^B}(\cdot|\mathbf{c}') \,\Big\|\, \pi_{\theta^B}(\cdot|\mathbf{c}')\right) + \\ &\alpha_3 \mathcal{D}_{KL}\left(\pi_{\phi^B}(\cdot|\mathbf{m}) \,\Big\|\, \pi_{\psi^B}(\cdot|\mathbf{o})\right)\Big]\end{aligned}}_{\text{Interchangeability loss}}.$$

where $\mathbb{E}_{\mathcal{I}}[\cdot]$ be the expectation induced by the $r.v.s.$ $\mathbf{m} \sim \pi_{\theta^A}(\cdot|c)$, $\mathbf{c}' \sim \pi_{\phi^B}(\cdot|m)$, $\mathbf{s} \sim \mu$, $\mathbf{o} = f_A(\mathbf{s})$, $\mathbf{o} \to \mathbf{c}$ and $\mathbb{E}_{\mathcal{I}_t}[\cdot]$ be the expectation induced by the $r.v.s.$

The goal is to find the optimal parameters θ^*, ϕ^*, ψ^* such that:

$$\theta^*, \phi^*, \psi^* = \arg\max_{\theta,\phi,\psi} J(\theta, \phi, \psi)$$

4.1 Reward Function

We follow a reward mechanism that balances exploration, cooperation, synchronization, accuracy, and efficiency in communication. In the Pathfinder game, which entails multi-step communication, agents are required to focus on delayed rewards. Agents are trained with shared rewards, fostering cooperation and shared language formation. To encourage agent exploration, we offer partial and complete rewards, motivating the agent to try different approaches and adapt themselves to make informed decisions during training. Both agents receive a partial reward if the listener infers the right region where the topic vertex is located but fails to identify the topic vertex. This acknowledges the successful transmission of relevant information without complete understanding. A full reward is given if the listener can accurately and unambiguously infer the exact topic vertex from the communicated information signifying a high level of successful communication and concept selection.

$$\begin{aligned}\mathbf{r} =& \mathbf{r}[1] + I(v_1 = v_1')\mathbf{r}[2] + \dots \\ &+ I(v_1 = v_1', v_2 = v_2' \dots v_{h-1} = v_{h-1}')\mathbf{r}[h],\end{aligned}$$

where $\mathbf{r}[j]$ is the reward associated with the j^{th} phrase which is defined as follows:

$$\mathbf{r}[j] = \begin{cases} \zeta_1 \ (\in \mathbb{R}), \text{ if } \ v_j = v_j', \\ \zeta_2 \ (\in \mathbb{R} \wedge \zeta_2 < \zeta_1), \text{ if } \ j = h \\ \qquad\qquad \text{and } C_z(v_j) \cap C_z(v_j') \neq \emptyset, \\ \zeta_3 \ (\zeta_3 \leq 0 \wedge \zeta_3 < \zeta_2), \text{ otherwise.} \end{cases}$$

A penalty is given if communication fails in order to discourage the respective concept-vocabulary mapping and to prevent incorrect or ineffective communication choices. The concept-selection module of the speaker seeks to select the optional spatial description to refer to the topic vertex by deactivating redundant concepts. The mechanism aims to ensure that the sentence corresponding to the generated spatial description is of optimal length to convey the intended meaning effectively. To support optimal word-order selection, we penalize the speaker for choosing a suboptimal sequence of concepts. In cases where a concept is deactivated, the agent chooses to remain silent at that particular instant of the corresponding generated message. To enable this, the utterance module chooses a NULL utterance $\Gamma(\square)$ to indicate silence. The concept of $\perp$ utterance is significant since we do not explicitly impose it a priori; rather, it is learned through interactions. In order to promote consistency and coherence in the use of the $\Gamma(\square)$ utterance across different word categories in a sentence, we employ a strategy to positively reward $\mathbf{r}'$ the speaker for the reuse of the same word for the $\square$ concept, irrespective of its temporal position in a phrase. This reward system encourages the emergence of a common word for the $\square$ concept across different contexts, regardless of its position in the message $\mathbf{u}_k$.

$$\mathbf{r}_t' = \begin{cases} \zeta_1' \quad (\in \mathbb{R}), \text{ if } |\{\Gamma(a) | a \in \mathbf{u}_t \wedge a = \perp\}| = 1, \\ \zeta_2' \quad (\zeta_2' < \zeta_1'), \quad \text{otherwise.} \end{cases}$$

4.2 Iterated Learning

In this paper, we consider an iterated learning framework to study the propagation and evolution of recursive language through generations, where the language has to pass through the bottleneck of cultural transmission. Iterated learning [24] is a framework to study emergent language structures by simulating cultural evolution. Iterated learning can be explained as the inductive process by which a certain behaviour develops in one individual through finite realizations of the same behaviour in another individual who acquired that behaviour in a similar manner. The iterated learning model is used to simulate the language evolution process based on the idea that the simulated language must be learned by new speakers at each generation from finite samples while also being used for communication [24, 25]. The language structure evolves through continuous and iterated learning due to inter-generational language discrepancies that arise when language is passed down through the bottleneck of inter-generational language transmission. In the iterated learning model (Figure 3), agents of generation

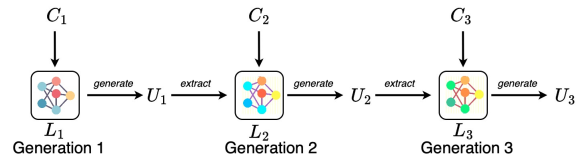

Figure 3: An agent from generation i, utilizing language L_i, produces utterances U_i to convey concepts C_i to the next generation, $i+1$. The agents of generation $i+1$ then derive their language L_{i+1} from these utterances, perpetuating the iterative refinement and transmission of language over generations.

$i+1$ update their language L_{i+1} by learning from the utterances U_i produced by the previous generation i. This iterative process drives the gradual evolution of linguistic structures across generations. As language is passed down, recurring patterns stabilize, leading to the emergence of systematic communication strategies.

5 EXPERIMENTS & DISCUSSIONS

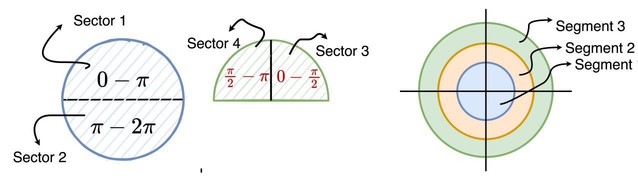

Figure 4: Concept space and the corresponding conceptualization of vertices.

In our experiments, we start with random initial values of the hyperparameters (learning rate, batch size, and regularization length). These hyperparameters are fine-tuned by a continuous process of rigorously observing the experiment results. We consider a graph with 6 vertices setup for our 2D grid-world environment. The source and target vertex pairs used for experiments have a maximum distance or depth of 2 between them, which indicates that either the source and target vertex are neighbours or they have one intermediate vertex between them. Initial experiments start with two homogeneous agents who can be seen as the adult population. This initial population of agents is involved in playing the PathFinder game and developing the initial communication protocol. Each dialogue consists of 100 conversations (batch size). The interchangeability property is explored by switching the roles of the speaker and listener every 500 epoch. The speaking and listening module within the agent's architecture utilizes an LSTM cell. Speaker's observation consists of coordinates of source location s, target location t. For having a local understanding of the whole environment, the observation vector also contains the angle and Euclidean distance of each vertex in the graph $w.r.t.$ source vertex. The lexis size ($|\Xi|$) is taken as 25. The observation vector o_t by the speaker agent is transformed into a feature vector ($\in \mathbb{R}^{25}$) by passing it through a fully connected neural network. This feature vector forms the hidden input of the concept selection module (ψ) whose hidden size is also taken as 25. The speaking and listening modules are implemented as a single-layer LSTM cell with a hidden size of 250. The LSTM networks output the sequence of words (for the speaker) or concepts (for the listener) with a maximum length of 3. For the Gumbel-Softmax-based continuous relaxation within the concept selection module, the temperature parameter τ is set to 0.5. Gradients originating from all modules are clipped with a maximum value of 50. Additionally, successful communication rewards both the speaker and listener with 100, and partial success merits a reward of 50. The concept space C consists of 4 sectors, 3 segments, and 4 colors. The concept space C is illustrated in Figure 4. Since there are overlapping sectors (Sectors 1, 3, and 4) we have a poverty of stimulus situation. We encode concepts using integers, where 0 is used for the null concept $\square$, $\{1, 2, 3\}$ represent segments belonging to $\mathcal{H}$, $\{4, 5, 6, 7\}$ for sectors $\mathcal{W}$ and $\{8, 9, 10, 11\}$ for colors $\mathcal{B}$ respectively.

During every conversation in a dialogue, a random vertex (except the source vertex) is chosen as the topic vertex. In this paper, we consider two-timescale networks [8] to obtain synchronized convergence, where the utterance network is calibrated using a faster timescale compared to the conception selection network. In this approach, the concept selection network can be considered to be pseudo-stationary, while the utterance network converges with respect to the stationary values of the concept selection network, and this cycle repeats itself in the long run. To achieve this, we employ the vanilla stochastic gradient algorithm with learning rates of the respective networks differing by order of magnitude.

In our experiments, we explicitly test the emergence, stability, and adaptability of recursive language in settings that mimic increasing environmental and architectural complexity. By exploring key language-shaping phenomena, such as dynamic environment complexity, role interchangeability, iterated learning, and an expanding lexis size, we provide strong evidence that recursive language not only emerges but thrives, even when faced with real-world challenges. These controlled yet demanding conditions reveal how recursive language can converge and remain stable in fluctuating environments.

5.1 Groundedness and Interchangeability

During the initial phase, agents' vocabulary mappings are randomly generated, allowing for exploration that drives the evolution of coherent language within a limited number of dialogues. Coherence in this context refers to the consistent and meaningful use of language, where words and expressions convey clear and shared meanings. This coherence is evidenced by the convergence of loss functions and the maximization of average rewards, as depicted in Figures 5 and 6. To improve communication, we adjust policy parameters using policy gradient methods over a non-differentiable channel between speakers and listeners. Positive rewards, indicating successful communication, are rare but crucial for reinforcing effective strategies. However, random neural network initialization and topic pair distribution may hinder this reinforcement, impacting the system's behavior.

Furthermore, the homogeneous agent architectures facilitate role-switching between agents, ensuring flexibility in communication. By incorporating a component for learning reverse mappings in the cost function, agents can sustain learned mappings during continuous language evolution. This is evident when agents switch roles periodically, minimizing the interchangeability loss ($\mathcal{L}_2$) in the cost function. Furthermore, the continuity of learning is demonstrated in Figure 8, where dips in success ratio (attributed to role switching) gradually diminish, ensuring near-seamless progress.

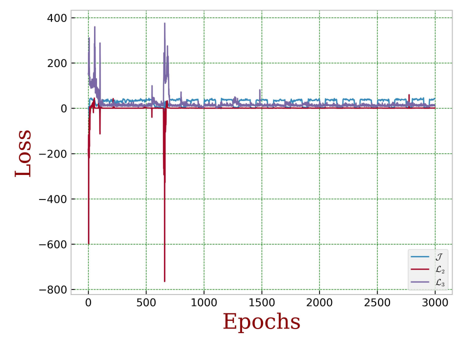

Figure 5: Evolution of the total cost function, interchangeability loss, and description length over time.

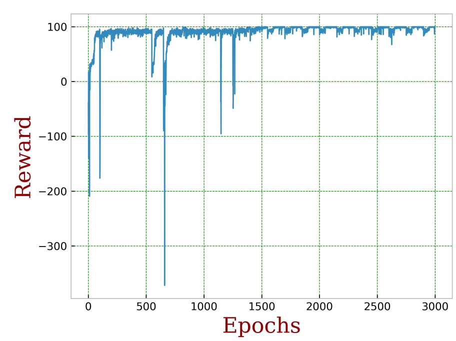

Figure 6: The mean reward achieved across interactions within a single dialogue.

5.2 Complex Goal Identification

During conversation, the speaker can either choose the depth-0 or depth-1 vertex. After choosing the topic vertex, both agents engage in a round of dialogue where the length of the message uttered by the speaker depends on the intermediate vertex between the source and target vertices. Initially, agents successfully communicate goals with a path length of 1. As shown in Figure 10, grounding emerges in the environment as agents are able to comprehend larger path length-2 goals. This ability stems from the recursive structure inherent in the language, enabling speakers to embed messages within a single sequence. The message conveyed for path length 1 serves as the foundation for achieving goals involving intermediate vertices with greater path lengths. Agents adhere to a recursive structure in their messages, as described in Section 3.1. For depth-1 samples (denoted by $P = \{s = v_0, v_1, t = v_2\}$), the message m combines the messages associated with the individual paths (v_0, v_1) and (v_1, v_2). In these cases, the agent simultaneously produces the sequence of messages $m = m_1, m_2$ simultaneously, and the listener unfolds the message using the recursive structure to navigate to the target vertex t.

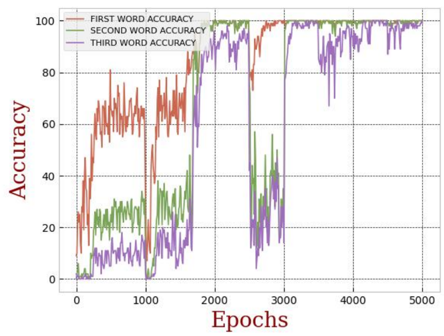

Figure 7: First, second, and third word accuracy.

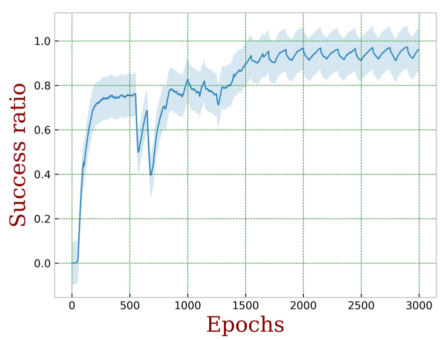

Figure 8: Ratio of correct target vertex prediction *i.e.*, **listener was able to navigate to target vertex.**

5.3 Cultural Evolution and Adaptation

We conducted experiments to explore the propagation of recursive language across successive generations. New agents, initially unfamiliar with the language, were introduced into the environment and assigned the role of listener for a certain period before being assigned the role of speaker. Two agents were introduced at different intervals (one at dialogue 5000 and the other at dialogue 10000) to observe how the language structure propagates. Throughout the training process, we observed significant fluctuations in success ratio and reward, as depicted in Figures 8 and 6, respectively. Notably, there are notable drops and subsequent rises in these metrics. By epoch 12000, all agents, including the newcomers, had converged to almost identical language structures as the original parents, as evidenced by the similarities in the language heatmaps shown in Figures 9 and 11.

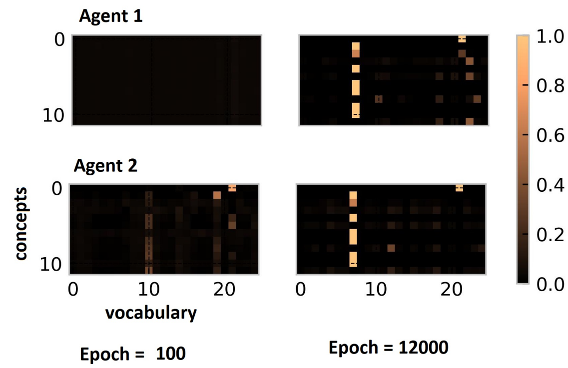

Figure 9: Consolidation of vocabulary symbols across concepts

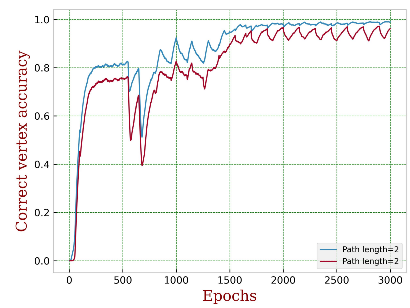

Figure 10: The accuracy for conversation between agents for goal vertex with distance 1 and 2

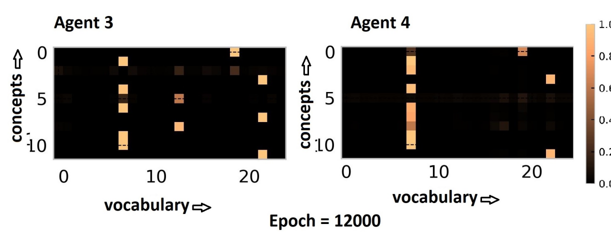

Figure 11: Consolidation of vocabulary symbols among child agents across concepts

Limitations: In some trials ($< 5\%$), successful behavior is not observed, primarily due to the lack of positive rewards resulting from random initialization of neural network weights and the distribution bias of the source-topic vertices pair chosen for conversations. Additionally, our study focuses on recursive language up to depth-1, which can be further scaled to explore deeper levels of recursion.

6 CONCLUSION

In this paper, we demonstrate the emergence of grounded recursive language, where longer messages are communicated by breaking them down into comprehensible parts. This approach ensures that all parts of lengthy messages are interconnected through shared contexts. We observed that agents learn to consistently use common symbols for identical concepts within messages. Our findings underscore the enduring nature of recursive structures, indicating their transmission across generations of communicators and emphasizing their fundamental role in shaping effective communication dynamics over time.

REFERENCES

[1] Jacob Andreas. 2018. Measuring Compositionality in Representation Learning. In *International Conference on Learning Representations*.

[2] Robert L Arrington. 1954. Ludwig Wittgenstein: Philosophical Investigations. In *Central Works of Philosophy v4*. Routledge, 257–279.

[3] Marco Baroni. 2020. Linguistic generalization and compositionality in modern artificial neural networks. *Philosophical Transactions of the Royal Society B* 375, 1791 (2020), 20190307.

[4] Ramon Ferrer I Cancho and Ricard V Solé. 2003. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences* 100, 3 (2003), 788–791.

[5] 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*. Association for Computational Linguistics, Online, 4427–4442. https://doi.org/10.18653/v1/2020.acl-main.407

[6] Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. 2019. Anti-efficient encoding in emergent communication. *Advances in Neural Information Processing Systems* 32 (2019).

[7] 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*.

[8] Wesley Chung, Somjit Nath, Ajin Joseph, and Martha White. 2018. Two-timescale networks for nonlinear value function approximation. In *International conference on learning representations*.

[9] Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R McKee, Joel Z Leibo, Kate Larson, and Thore Graepel. 2020. Open problems in cooperative AI. *arXiv preprint arXiv:2012.08630* (2020).

[10] Gautier Dagan, Dieuwke Hupkes, and Elia Bruni. 2020. Co-evolution of language and agents in referential games. *arXiv preprint arXiv:2001.03361* (2020).

[11] LEWIS David. 1969. Convention: a philosophical study.

[12] Ferdinand de Saussure. 2011. *Course in general linguistics*. Columbia University Press.

[13] Tom Eccles, Yoram Bachrach, Guy Lever, Angeliki Lazaridou, and Thore Graepel. 2019. Biases for emergent communication in multi-agent reinforcement learning. *Advances in neural information processing systems* 32 (2019).

[14] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems* 29 (2016).

[15] Emilia Garcia-Casademont and Luc Steels. 2016. Insight Grammar Learning. *Journal of Cognitive Science* 17, 1 (2016).

[16] Miquel Cornudella Gaya, Thierry Poibeau, and Remi van Trijp. 2016. The role of intrinsic motivation in artificial language emergence: a case study on colour. In *26th International Conference on Computational Linguistics (COLING 2016)*. 1646–1656.

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

[18] Shubham Gupta, Rishi Hazra, and Ambedkar Dukkipati. 2020. Networked Multi-Agent Reinforcement Learning with Emergent Communication. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (Auckland, New Zealand) *(AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1858–1860.

[19] Marc D Hauser, Noam Chomsky, and W Tecumseh Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *science* 298, 5598 (2002), 1569–1579.

[20] Serhii Havrylov and Ivan Titov. 2017. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. *Advances in neural information processing systems* 30 (2017).

[21] Eric Jang, Shixiang Gu, and Ben Poole. 2016. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144* (2016).

[22] Ivana Kajić, Eser Aygün, and Doina Precup. 2020. Learning to cooperate: Emergent communication in multi-agent navigation. *arXiv preprint arXiv:2004.01097* (2020).

[23] Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. 2020. Entropy minimization in emergent languages. In *International Conference on Machine Learning*. PMLR, 5220–5230.

[24] Simon Kirby, Tom Griffiths, and Kenny Smith. 2014. Iterated learning and the evolution of language. *Current opinion in neurobiology* 28 (2014), 108–114.

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

[26] Satwik Kottur, José MF Moura, Stefan Lee, and Dhruv Batra. 2017. Natural language does not emerge'naturally'in multi-agent dialog. *arXiv preprint arXiv:1706.08502* (2017).

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

[28] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. 2018. Emergence of linguistic communication from referential games with symbolic and pixel input. *arXiv preprint arXiv:1804.03984* (2018).

[29] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-Agent Cooperation and the Emergence of (Natural) Language. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

[30] Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. 2020. Multi-agent Communication meets Natural Language: Synergies between Functional and Structural Language Learning. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 7663–7674. https://doi.org/10.18653/v1/2020.acl-main.685

[31] David Lewis. 2008. *Convention: A philosophical study*. John Wiley & Sons.

[32] Fushan Li and Michael Bowling. 2019. Ease-of-teaching and language structure from emergent communication. *Advances in neural information processing systems* 32 (2019).

[33] Toru Lin, Jacob Huh, Christopher Stauffer, Ser Nam Lim, and Phillip Isola. 2021. Learning to ground multi-agent communication with autoencoders. *Advances in Neural Information Processing Systems* 34 (2021), 15230–15242.

[34] Michael L Littman. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*. Elsevier, 157–163.

[35] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. 2016. The concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712* (2016).

[36] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).

[37] Igor Mordatch and Pieter Abbeel. 2018. Emergence of grounded compositional language in multi-agent populations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[38] Andrew Nevins, David Pesetsky, and Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* (2009), 355–404.

[39] Amy Perfors, Josh Tenenbaum, Edward Gibson, and Terry Regier. 2010. How recursive is language? A Bayesian exploration. *Recursion and human language* (2010), 159–175.

[40] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[41] Shuwen Qiu, Sirui Xie, Lifeng Fan, Tao Gao, Jungseock Joo, Song-Chun Zhu, and Yixin Zhu. 2022. Emergent graphical conventions in a visual communication game. *Advances in Neural Information Processing Systems* 35 (2022), 13119–13131.

[42] Willard Van Orman Quine. 1960. Word and object MIT press. *Cambridge MA* (1960).

[43] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B Cohen, and Simon Kirby. 2020. Compositional languages emerge in a neural iterated learning model. *arXiv preprint arXiv:2002.01365* (2020).

[44] Cinjon Resnick, Abhinav Gupta, Jakob Foerster, Andrew M Dai, and Kyunghyun Cho. 2020. Capacity, Bandwidth, and Compositionality in Emergent Language Learning. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1125–1133.

[45] Mathieu Rita, Rahma Chaabouni, and Emmanuel Dupoux. 2020. " LazImpa": Lazy and Impatient neural agents learn to communicate efficiently. *arXiv preprint arXiv:2010.01878* (2020).

[46] Mathieu Rita, Florian Strub, Jean-Bastien Grill, Olivier Pietquin, and Emmanuel Dupoux. 2022. On the role of population heterogeneity in emergent communication. *arXiv preprint arXiv:2204.12982* (2022).

[47] Lloyd S Shapley. 1953. Stochastic games. *Proceedings of the national academy of sciences* 39, 10 (1953), 1095–1100.

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

[49] Luc Steels. 2003. Evolving grounded communication for robots. *Trends in cognitive sciences* 7, 7 (2003), 308–312.

[50] Luc Steels. 2012. Grounding language through evolutionary language games. In *Language grounding in robots*. Springer, 1–22.

[51] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. *Advances in neural information processing systems* 29

(2016).
[52] Ming Tan. 1993. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*. 330–337.
[53] Mycal Tucker, Roger Levy, Julie A Shah, and Noga Zaslavsky. 2022. Trading off utility, informativeness, and complexity in emergent communication. *Advances in neural information processing systems* 35 (2022), 22214–22228.
[54] Kyle Wagner, James A Reggia, Juan Uriagereka, and Gerald S Wilkinson. 2003. Progress in the simulation of emergent communication and language. *Adaptive Behavior* 11, 1 (2003), 37–69.
[55] Sida I. Wang, Percy Liang, and Christopher D. Manning. 2016. Learning Language Games through Interaction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 2368–2378. https://doi.org/10.18653/v1/P16-1224
[56] Ludwig Wittgenstein. 1954. *Philosophical investigations.* John Wiley & Sons.
[57] George Kingsley Zipf. 2016. *Human behavior and the principle of least effort: An introduction to human ecology.* Ravenio Books.