Reinforcement Learning in Language Evolution

Martin Bachwerk
Computational Linguistics Group,
School of Computer Science and Statistics,
Trinity College, Dublin, Ireland

Abstract. Experiments on the evolution of language conducted with two success-based implementations of Herrnstein reinforcement learning exhibit significantly higher performance rates for populations of agents who employed a moderate requirement for minimum success to decide if a lexical item should be positively or negatively reinforced. Furthermore, a dynamic reinforcement model based on the actual success of an interaction significantly outperforms a model with fixed-value reinforcements.

1 Introduction

Looking at the literature on the evolution of language, one would discover an abundance of more or less convincing biological [16, 9], social [8, 7] and linguistic [5, 6, 11] theories on how our distant predecessors could have invented such a complex communicative tool as human language. Admittedly, many of the arguments presented in such manuscripts are extremely plausible and some of them must certainly be very close to reality. However, even the most trusting scholar would very quickly notice a certain lack of specificity when going through the research published in the area of evolutionary linguistics. For example, numerous theories claim to explain how and why the first symbolic, learned words were utilized, yet the authors seem to always sidestep the challenges involved in learning the very first words and employing these as worthwhile communication tools.

– How does the hearer know exactly enough what the speaker is talking about?
– What happens if two interlocutors seem to know the words, but in reality are talking about two completely different things (or vice versa)?

These and many other questions outlined for example in Quine’s indeterminacy of translation thesis (cf. [17]) remain largely unanswered by the existing literature on the emergence of language, presumably (and understandably) due to a complete lack of linguistic evidence from the evolutionary process itself. The advance of computing technology in recent decades has now made it possible for researchers to tackle issues as complex as the ones outlined above, in particular with the help of mathematical models and computer simulations.

The goal of this paper is accordingly to provide a computational model, if not unequivocally answering the posed questions, at the very least providing an
empirical basis and some motivation for future research in the field. In order to place the presented model in appropriate context, Section 2 will outline the general idea of the signaling game – a class of game-theoretic models that are applicable to the question of language evolution. Section 3 will give a short overview of the arguably most prominent computational study of the emergence of lexicon that also implements some aspects of the signaling game. Section 4 will then present an alternative computational model for performing simulations on the evolution of language in the context of the signaling game, along with its most recent developments. In the following Section 6, some experimental results of simulations performed with the help of the model will be discussed. Finally, Section 7 will round off the paper with some conclusions and considerations for future work.

2 The Signaling Game

The signaling game introduced by [15] represents a class of games that can be employed for the investigation of evolution of a variety of signaling systems, including the human language (see [13] for an extensive review). In its original form, the signaling game scenario includes two equiprobable world states, two agents, two signals and two actions. The sender agent has alone the access to the world state, which he may or may not use when producing a signal. The hearer agent observes the signal and may or may not consider it when deciding on an action to perform. The essence of the game lies in the definition of the payoffs for the possible state-action combinations. In particular, of the two actions that are present in the model, exactly one is appropriate for any particular world state. If the hearer performs the action that is appropriate for the world state that the agents find themselves in, he will obtain a payoff. Given that the signaling game is meant to simulate cooperative behaviour, the speaker also receives a payoff if the hearer performs the correct action for the given state. The different costs that may incur during signal production or execution of actions will not be considered in this paper.

The goal of the signaling game is for the two agents to arrange their production and response rules in a way that a maximum payoff is obtained at all times by both agents, i.e. reaching an optimal Nash equilibrium, also called a signaling system. Apart from the optimal equilibrium, there are also other Nash equilibria that do not yield a perfect payoff, e.g. a pooling equilibrium, in which agents always perform the same action regardless of the heard signal. When applying a learning technique to the signaling game, it is usually desirable to avoid getting stuck in such (partial) pooling equilibria. It has been shown that for the simple 2-state/2-term version of the game, Herrnstein reinforcement learning (cf. [4]), according to which agents update their propensity for a particular signal or a response action in direct proportion to the payoff it yielded in a signaling game, will always converge on a signaling system (cf. [1]). However, if the complexity of the game were increased, further modifications to the learning strategy would be required in order to make sure that agents do not get stuck in (partial) pooling
equilibria. A possible modification is to introduce punishments to the updating rule, which would reduce the propensities for a particular action if it has not resulted in any (or sufficient) payoff (cf. [3]).

3 The Guessing Game

The guessing game developed by [19] is a branch of the so called language games which were designed with the goal of simulating the evolution of language, predominantly with the help of embodied agents. An interaction scenario in the guessing game consists of two agents, a context, which is usually a set of objects, and a topic, selected by the speaker from the set of context objects without the knowledge of the hearer. The task of the hearer is then to correctly discriminate the topic with the help of available categorization mechanisms. The agents in the guessing game implement a learning strategy that is a variant of Herrnstein reinforcement learning which not only increases/decreases the weights of the rules used in the current interaction, but, in case of success, also inhibits the weights of all competing rules. While this so called lateral inhibition approach has shown to yield near-optimal results in a number of simulations (cf. [19, 20]), this paper will concentrate only on act-based learning models, i.e. those where only the rules employed in a given action can be updated thereafter. A further issue with the implementation of the guessing game is that in order to determine the success of an interaction, the hearer is presented with the exact meaning of the encountered utterance after his attempt to guess it. Unfortunately, this kind of explicit feedback does not reflect evolutionary reality and is in effect an escape hatch which allows to sidestep Quine’s indeterminacy of meaning issues. An alternative way of approximating success which has been employed in the simulations presented in this study is given in the following section.

4 Language Evolution Workbench

As an alternative to the approach presented above, the Language Evolution Workbench (LEW) was introduced by [23, 2] with the aim of being able to perform a wide range of experiments (the workbench currently has over 20 adjustable parameters) on the evolution of language without having to make any unnecessarily strong prior assumptions. The few assumptions that are made can be considered as widely accepted (according to [14] and [21] among others) as the minimal prerequisites for the emergence of language. These include the ability to observe and individuate events, the ability to engage in a joint attention frame fixed on an occurring event, and the ability to interact by constructing words and utterances from abstract symbols and transmitting these

1The workbench will be made available to the general public in the near future under http://www.scss.tcd.ie/disciplines/intelligent_systems/clg/clg_web/LEW.
to one’s interlocutor. The remainder of this section provides a more detailed technical introduction into the different aspects of the LEW model. Since the basic structure of the model has not changed since [2], a reader who is familiar with that work can safely skip on to Section 5, which describes the specifics of this particular experiment.

**Agents, Entities and Events** Agents in the LEW are non-physical entities (see [19] and further works by Luc Steels and his colleagues for embodied implementations) and are not specialized to the question of language evolution. What characterizes every agent in the LEW is solely a knowledge base and a lexicon. The knowledge base consists of all experienced events in the order in which an agent encountered these and without any property-specific organization of meaning, i.e. conceptualization. The lexicon is represented as a set of $<\text{Meaning, Form, Weight}>$ tuples, which are also referred to as lexical items or mappings. In such a meaning-form mapping, the *Meaning* is (a part of) some event that was encountered by the agent and the *Form* is (a part of) some utterance that was either produced or heard by the agent in relation to the event. The *Weight* is an indicator of confidence in the mapping, and, depending on the experimental setup, is adjusted according to one of the equations provided in Section 5. In addition, mapping weights can also decay over time, if the forgetting parameter is enabled.

Events that occur in the LEW are generated by selecting one of the predefined event types and filling it with arguments. Hereby, an *event type* simply specifies the ordering and the types of arguments that are permitted for event instances of the given type, with seven different types available in the model (corresponding argument types are noted in parentheses):

- $R^1$ [animate]
- $R^2$ [animate, animate]
- $R^3$ [animate, unconstrained, animate]
- $R^4$ [animate, animate, inanimate]
- $R^5$ [animate, event]
- $R^6$ [unconstrained]
- $R^7$ [animate, unconstrained, event]

Whenever an event needs to be generated, an event is selected based on a predefined Zipfian distribution. Afterwards, a random (but type-appropriate) argument is selected for each slot in the event. As can be seen from the above

---

2By default, the hearers are not even assumed to know the word boundaries of an encountered utterance. However, simulations with so called synchronized transmission have been also performed and reported on by [22].

3A concept-based world model is undoubtedly a very prominent feature that is missing from the agents in the LEW model. However, it has been shown that humans are not the only species that are capable of organizing the world in concepts (cf. [10]), yet they are the only ones that have a communication system as complex as language. Further proof for the secondary role of conceptualization is that, in the LEW, linguistic conventions emerge in simple agents even without concepts.
listing of event types, an argument of an event can either be an entity or another event, in which case the selection process is repeated recursively, thus resulting in an unbounded meaning space based on a finite number of both entities and event types. Entities are represented as propertyless atoms, and an arbitrary number of these can be experimented with. However, in the current work, we define entities in terms of sorts, whereby two sorts are distinguished, namely animates and inanimates. To provide the reader with a more concrete example, let us take a look at a sample event [ihdixos animate davr animate]. In this example, ihdixos is the event atom corresponding to the event type with an animate first argument and an event second argument (R1), whereby davr is the event atom corresponding to the recursively selected event type with one animate argument (R1).

Interactions Building on the traditions of computer simulations of language evolution, the LEW simulates interactions between agents. Every interaction in the LEW occurs between two randomly chosen agents, a speaker and a hearer, whereby an agent can also end up talking to himself if he gets picked as the hearer too (language is meant for thinking as well as communicating). More specifically, the interactions following the following scenario:

**Step 1:** An event is constructed as described above and fully ‘displayed’ to both agents.

**Step 2:** The speaker builds up his perspective on the event by arranging the event into segments. The model builds in no constraints on this process, so the an event like [ihdixos animate davr animate] could be arranged into anything between one combined meaning unit [[ihdixos animate davr animate]] to four individual meaning units [[ihdixos animate] [davr animate]].

**Step 3:** The speaker assigns a lexical form, i.e. a word, to every meaning unit from the previous step, combines the words into one continuous utterance and transmits it to the hearer. The assignment is achieved by either looking up an existing form for a meaning unit in one’s lexicon, or, failing to find one, by inventing a new form. A sample utterance for the four-unit segmentation of the above event could be something like “ii-j i-w p-t k-b”, where each word corresponds to precisely one of the meaning units, in the same order.

**Step 4:** The hearer receives the utterance as a continuous stream of uninterrupted sound and first of all needs to segment the utterance into individual words, or decide that he regards the utterance as a single, and possibly very long word. This is referred to in the LEW as asynchronized transmission. An experiment with synchronized transmission, i.e. where the hearer knows the exact word boundaries have been also performed and reported on by [22].

4If multiple synonymous meaning-form mappings exist in an agent’s lexicon, then a form is randomly picked based on a probability distribution computed from the corresponding mapping weights (see below for more on weights).

5This is referred to in the LEW as asynchronized transmission. An experiment with synchronized transmission, i.e. where the hearer knows the exact word boundaries have been also performed and reported on by [22].
Step 5: The hearer assigns a meaning to every word he thinks to have encountered in the utterance. This is done in a similar way to how the speaker composes an utterance. First of all, the hearer looks in his lexicon for an existing meaning that is stored in relation to the heard form. Failing to find one, the hearer randomly assigns some segment of the event as the meaning, whereby his perspective on the event, i.e. its segmentation into meaning units, is fully independent of the speaker’s perspective (cf. step 3).

Step 6: In order for the agents to learn, an update is performed for every meaning-form mapping that they employed in an interaction. This update mechanism is described in more detail in Section 5.

Understanding and Success While much of the effort in the LEW has been put into the avoidance of direct meaning-form transfer (i.e. telepathy), it is nevertheless possible to observe the levels of understanding between two interacting agents, e.g. by looking at the precision and recall of the hearer’s understanding of a speaker’s utterance. To summarize both precision and recall in one measure, the analysis will make use of a statistical measure called the F1 score that represents the harmonic mean between precision and recall, defined as follows:

\[
F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

When making external measurements, it has to be noted that one need not necessarily look at direct form-to-meaning understanding as it can be safely assumed that, in the pessimistic setting of agents individuating events differently, they will scarcely share any mappings. However, a certain level of understanding can still be accredited to the interlocutors if they have somehow reached the same meaning, even though they expressed it differently. Pragmatically, there can be some level of success if distinct symbol-meaning mappings take agents to at least some of the same meanings. If an agent wants salt, asks for salsa, and obtains salt, then some amount of communicative success has happened.

Furthermore, if we assume that communication involved out of the need for cooperation, and that successful communication will lead to clearly perceivable payoffs during such cooperation, then it is acceptable to approximate the success of a potential action over the understanding rate of a given interaction. In effect, this approach lets the agents know how appropriate were the actions that they performed in response to an interaction, with the assumption that action success is directly dependent of communicative success. Note that this type of implicit feedback is provided to the agents without letting them know precisely what part of the utterance it was that they succeeded in or failed to understand, thus still avoiding any kind of telepathy-like meaning transmission. In the presented study, the approximated measure of success has been utilised directly in the two implemented learning strategies, which are described in further detail below.

This approach follows [18] in saying that complete understanding is not a requirement for the emergence of conventionalized communicative mechanisms.

---

6 If an agent wants salt, asks for salsa, and obtains salt, then some amount of communicative success has happened.

7 This approach follows [18] in saying that complete understanding is not a requirement for the emergence of conventionalized communicative mechanisms.
5 Experiment Design

Prior to the current study, all simulations of the LEW implemented a very primitive learning strategy that essentially reinforced all meaning-form mappings utilized to either construct or decode an utterance without any consideration for their success. Based on the notation from [4], the updating rule for an agent’s propensity \( q_i(t) \) to utilize a lexical mapping \( i \) at time \( t \) in this learning strategy could be defined for every lexical mapping \( i \) employed in an utterance as:\(^8,9\)

\[
q_i(t + 1) = \begin{cases} 
q_i(t) + \pi(t) & \text{if mapping } i \text{ was utilized} \\
q_i(t) & \text{otherwise} 
\end{cases}
\]  

Where \( \pi(t) \) represents the payoff of the interaction performed between two agents at time \( t \) and was simply fixed as:

\[
\pi(t) = 1
\]  

As can be quite clearly seen, the above equations do not reflect the success of an employed strategy. Consequently, there is also no way of introducing punishments for unsuccessful strategies, which have proven to be essential if one wishes to reach any kind of near-optimal solution in the more complex signaling games (cf. [4]), which the LEW clearly represents. As an improvement to the learning strategy outlined above, this study proposes to approximate the payoff of an interaction based on its communicative success \( s(t) \), which can range from 0 (no understanding) to 1 (full understanding). In order to do so, an additional parameter that defines the level of minimum success \( s_{\text{min}} \) was introduced to the LEW. This parameter allows for a more flexible payoff definition, which incorporates both reinforcement and punishment of a strategy:

\[
\pi(t) = \begin{cases} 
1 & \text{if } s(t) \geq s_{\text{min}} \\
-1 & \text{otherwise} 
\end{cases}
\]  

While the payoff definition in equation (3) is surely an improvement on the completely success-agnostic reinforcement employed previously as described in equations (1,2), it still fixed at a certain value and is not dependant on the actual distance between the level of minimum success and the actual success rate. This shortcoming can be solved by a more dynamic payoff definition:\(^10\)

\[
\pi(t) = s(t) - s_{\text{min}}
\]  

\(^8\)Since action success in the LEW is approximated over communicative understanding, no response rule needs to be defined here. Instead, much like the Saussureans in [12], agents make use of the same lexicon both during signal production and interpretation, meaning that the updating rule applies to both sides of an interaction.

\(^9\)In the LEW, \( q_i \) values represent the weights and thus the usage probabilities of competing meaning-form mappings an an agent’s lexicon. In general, \( q_i(0) \) can be considered to be equal to 0. Note that this has no effect on the invention of new forms.

\(^10\)Having \( s_{\text{min}} = 0 \) in this case would make the updating rule equivalent to its basic version in Herrnstein reinforcement learning which employs just the actual payoff value.
The goal of the current study is to observe how the two different payoff definitions fare in comparison to a population that employs no learning strategy at all, or one that follows the success-agnostic strategy described by equations (1,2).\textsuperscript{11} Furthermore, the design of the study begs the question if an optimal level of minimum success becomes apparent from the performed simulations. In total, experiments with seven different conditions have been performed within this study: one with no learning involved, three implementing the fixed-value and three implementing the dynamic payoff definition. In the latter two cases, the level of minimum success was varied between 0, 0.25 and 0.5.

For the purpose of evaluating the performance of the different learning strategies (i.e. combinations of payoff definitions and the level of minimum success), measurements of understanding success rates were taken throughout the simulations and compared. Additionally, a range of further qualitative measures of the system’s linguistic properties were analysed. Taken together, this should provide an indication of how well equipped and also how realistic the emergent signaling systems really are. The most significant of the observed measures were: lexicon size, which depicts the gross potential of a lexicon; lexicon use rate and lexicon precision, which describe its actual usefulness; as well as agent synonymy and homonymy, which indicate the coherence levels of a lexicon. Of the above, lexicon size, as well as synonymy and homonymy rates are typical measures that appear in most works that involve some sort of a communication system. The two other measures – lexicon use and precision – were introduced here to be able to quantify the amount and the significance of guessing involved in the LEW.\textsuperscript{12}

6 Results and Discussion

The results provided in this section are based on 600 simulation runs of each condition, with every run including a total of 5000 interactions between agents. While there may exist quite convincing arguments for the case that the simulations could have been performed with a larger number of interactions, it can be equally argued that in order for a complex evolutionary trait to establish itself in a population, it should have provided at least some value from the very early stages, otherwise it would have been quickly abandoned by its pioneers. However, one would not expect perfect communication to evolve within a couple of hundred interactions per agent (which could be a matter of a few days). Accordingly, the essential aspect here is that a certain social behaviour, i.e. a learning strategy should display sufficient potential in its early stages in order for it to become a reasonable long-term candidate for an evolutionary adaptation.

Looking at Figure 1(a), which depicts the average understanding rates over the most recent 500 interactions, it can be seen that all but one learning strategy

\textsuperscript{11}This strategy is quite clearly a special case of equation (3) where \( s_{\text{min}} = 0 \).

\textsuperscript{12}In effect, both these measures can be meaningfully applied to any system that involves lexical innovations. In systems without innovations, e.g. the Lewis signaling game, lexicon use will be always equal to 1 and lexicon precision – to the overall communicative success of the agents.
exhibit a certain amount of learning, resulting in their success rates rising significantly above random levels ($t \geq 49.23$, $p < 0.0001$ for each of these strategies). The only learning strategy that actually falls below random guessing over time ($t = -3.16$, $p < 0.0016$) is the one employing a fixed-value payoff definition with minimum success level set to 0. What happens in this learning strategy is that every used mapping, regardless of its success or failure, is recorded in an agent’s lexicon, without ever being inhibited or forgotten. Such lexicon-hungry agents, whose lexicon size at the end of the simulations is on average between 2.33 and 8.86 times larger than that of others, as seen on Figure 1(b), become eventually incapable of either producing quality signals or reliably interpreting them.

The reasons for the agents with larger lexicons being at a disadvantage are among other things increased lexicon synonymy levels observed in such agents ($t \geq 24.72$, $p < 0.0001$ for every decreasing level of $s_{min}$ with fixed payoff; $t \geq 30.31$, $p < 0.0001$ for every decreasing level of $s_{min}$ with dynamic payoff) (cf. Figure 2(a)), a sparser distribution of lexical items among agents ($t \leq -55.97$, $p < 0.0001$ for every decreasing level of $s_{min}$ with fixed payoff; $t \leq -8.54$, $p < 0.0001$ for every decreasing level of $s_{min}$ with dynamic payoff) (cf. Figure 2(b)), as well as weaker dominant forms in the case where a meaning can be expressed by multiple synonyms for some of the learning strategies (this effect is not monotonic as the other two) (cf. Table 1).

While Figure 2(a) suggests that lexicon synonymy is a stably increasing phenomenon for all given learning strategies, albeit quite expectedly hovering at significantly lower levels for higher values of $s_{min}$, a very important observation to take from Table 1 is that for the dynamic payoff, the relative probability of one particular dominant form being used is significantly higher ($t = 63.61$, $p < 0.0001$), which should logically result in higher chances of two agents actually understanding each other, regardless of the overall amount of synonymy that still remains in their lexicons.
In order to determine which learning strategy actually performs best, one should take a further look at the agents’ lexicons, as these are solely responsible for the evolution of communicative success rates depicted in Figure 1(a) (the success of random guessing will be always the same, regardless of the learning strategy employed). Two questions seem to be of particular interest: how often is an agent’s lexicon actually used when trying to decode a signal and how good is the decoding precision of the lexicon in the cases where it was used. The answers to both these questions are presented in Figures 3(a) and 3(b) respectively.

The crucial aspect of Figure 3(a), which depicts the ratio of mappings that the hearer agents were able to find in their lexicon when attempting to interpret an utterance, is that lexicon use is constantly increasing for all learning strategies, presumably as a function of increasing agent lexicons, which is apparent from the fact that learning strategies with lower success requirements tend to end up having the highest lexicon use rates. However, lexicon use alone is only half of the solution, as exemplified by the very low (and the only decreasing) success rates of the learning strategy that employs a fixed-value payoff definition and \( s_{\text{min}} = 0 \), making it eventually perform worse than random guessing, as discussed previously. Consequently, the best strategy would be the one that results in a lexicon that can be employed as frequently as possible, yet is also sufficiently reliable. If one were to compare all of the lexicon-related data, then the strategy with the dynamic payoff definition and \( s_{\text{min}} = 0.25 \) should seemingly be the most optimal one. Perhaps not surprisingly, this strategy does indeed reach the best understanding rates overall (t value between 12.49 and 80.47, \( p<0.0001 \) when compared against other learning strategies), as seen from Figure 1(a).
7 Conclusions and Future Work

In conclusion, while more detailed models of language evolution like the guessing game or the LEW tend to become extremely complicated for a thorough formal analysis, they can nevertheless provide invaluable empirical insights into the mystery of human language evolution from a game-theoretical perspective. The current study attempts to prove this point by presenting a set of computer simulations that, with some extensive analysis, appear to reveal some valuable information about different potential learning strategies that our predecessors might have employed in order to learn the first language. In particular, the results suggest that employing a learning strategy with a success-based payoff definition has a significant effect on the reliability of the learned communication system, provided that the agents’ success expectations are set appropriately.

Further research in this area might benefit from looking more closely into the role that forgetting, especially in the form of temporal decay, might play on the elimination of unsuccessful, and thus rarely used lexical items. Another potential solution to synonymy elimination is lateral inhibition of competing meaning-form mappings on every successful use of a lexical item. Unfortunately though, it remains unclear how forgetting is actually represented in humans and thus further research is also necessary in the fields of neuroscience and psychology in order to evaluate the viability of either forgetting implementation in evolutionary terms.

References


13 Similarly to signaling games with a larger and/or open meaning and signal space, for which convergence theories do not yet exist.


