A New Approach to Exploring Language Emergence as Boundedly Optimal Control in the Face of Environmental and Cognitive Constraints

Jeshua Bratman[1] (jeshua@umich.edu)
Division of Computer Science and Engineering, University of Michigan, Ann Arbor, MI 48109

Michael Shvartsman[1] (mshvarts@umich.edu)
Richard L. Lewis (rickl@umich.edu)
Department of Psychology, University of Michigan, Ann Arbor, MI 48109

Satinder Singh (baveja@umich.edu)
Division of Computer Science and Engineering, University of Michigan, Ann Arbor, MI 48109

Abstract

Computational experiments have been used extensively to study language emergence by simulating the evolution of language over generations of interacting agents. Much of this work has focused on understanding the mechanisms of how language might have evolved. We propose a complementary approach helpful in understanding why specific properties of language might have emerged as an adaptive response to joint pressures from the environment and constraints on an agent's cognitive architecture. The approach suggests that linguistic systems can be described as boundedly optimal policies in multi-agent dynamic control problems defined by specific environments, agent computational structures, and task-oriented (vs. communication oriented) rewards. We illustrate the approach with a set of computational experiments.

Keywords: language emergence, bounded optimality, cognitive architecture, reinforcement learning, adaptive control

Introduction

The goal of this paper is to begin exploring a new approach to understanding the emergence of language. The primary scientific aim is understanding how pressures from the *environment* and constraints on the agent's *cognitive architecture* jointly lead to the emergence of specific properties of linguistic communication as optimal policies for obtaining well-defined long-term task- or environment-related reward.

Taking this perspective allows us to abstract away from the question of *how* language evolved and systematically explore constraints explaining *why* language appeared in the form that it has. We hypothesize that specific language-like properties (for instance, compositionality and systematic reliance on surface cues such as order) can in part be explained as bounded optimal solutions to control problems faced by computationally limited agents in environments exerting specific pressures. We propose investigating language through such environments in which we can formulate control problems for two or more bounded agents. If the optimal policies for these agents exhibit certain linguistic properties, then we can begin to define a mapping from the original pressures and agent constraints to the properties exhibited.

Finding solutions to these control problems computationally can be accomplished through various means such as reinforcement learning, game-theoretic analysis, or evolutionary algorithms. Thus, the approach allows us to step away from assumptions about specific mechanisms of learning or evolution, and focus on the joint relationship of agent structure and environment to derived linguistic systems. A feature of this approach that distinguishes it from related efforts is the focus on deriving control for internal cognitive processes and external actions generally rather than communication systems specifically, with communication processes emerging only if they are part of the optimal policy.

This paper proceeds as follows: first, we review related work on language emergence and discuss ways in which our approach complements this work. Next, we move to an example (the "Treasure Box Domain") designed to illustrate the approach by exploring constraints leading to the emergence of structured utterances — here the systematic use of serial order and allocation of lexical items to aspects of the environment. Finally, we show how this domain, and the approach in general, can be extended to investigate more sophisticated phenomena and propose future directions of inquiry.

Related Work

Research into the origins of language has a rich and controversial history. Chomsky addressed it in his early work on generative grammar, prompting a longstanding debate on the extent to which language is a biological adaptation arrived at via natural selection (Chomsky, 1968; Pinker & Bloom, 1990; for a more recent treatment, see Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005; Fitch, Hauser, & Chomsky, 2005; Jackendoff & Pinker, 2005). Chomsky's (Chomksy, 2010) own recent approach to the question attempts to minimize—in fact, nearly eliminate—the role of language-specific biological adaptation. A more recent line of research by Nowak and colleagues (Nowak, Krakauer, & Dress, 1999; Nowak & Krakauer, 1999; Nowak, Plotkin, & Jansen, 2000; Nowak, Komarova, & Niyogi, 2002), establishes a mathematical framework used to explore the evolution of language from the standpoint of computational learning theory and evolutionary game theory. This work also provides evidence for coding constraints that may have resulted in increased fitness for agents capable of multi-symbol utterances.

[1] The first two authors contributed equally to this paper.

Several recent computational experiments explore the notion that cultural adaptation and domain-general cognition may be sufficient for the emergence of language (Beckner et al., 2009, also see Christiansen & Chater, 2008; Steels, 1998; De Beule, 2008; Gong, Minett, Ke, Holland, & Wang, 2005). This work shows a number of features emerging from repeated interactions of pairs of computational agents in a population playing a language game. In a way, this work implicitly frames language emergence as a function of environment, agent, and learning mechanism. Our work attempts to remove the last of these and more explicitly address what aspects of environment and agent architecture are important—potentially leading to a more deeply explanatory account.

The questions we are interested in are in part orthogonal to these debates: we are not making claims about either domain-specificity or the mechanisms of learning or evolution, but rather the interplay of cognitive constraints and environmental pressures that lead to the emergence of particular language features as adaptive. By leaving the mechanism of adaptation unspecified, our approach is relevant to researchers working in both biological and cultural frameworks.

Our work also departs from the approaches above in that it does not create a pressure for language by explicitly rewarding cooperation or communication of a particular type. This approach considers communication not as an end-goal but rather as the means to obtain some primary reward such as sustenance, shelter or reproduction. This may give us a principled way to examine and sharpen what it is about language which directly contributes to effective behavior.

Environmental Pressures & Agent Constraints

Natural environments comprise extremely complicated sets of pressures acting on agents. A key part of the work in this approach is identifying tractable sets of specific pressures that are independently motivated by the study of the environments of early hominids or humans and that might plausibly be important in the emergence of language. It is not our intent in this initial exploration to undertake this identification systematically, but we propose here a few plausible candidates as starting points that suffice to illustrate the approach.

Many environments naturally limit agent's ability to observe and act. For example human beings can only manipulate small pieces of the natural world. Furthermore, knowledge and ability to act is not usually distributed uniformly among agents, making information sharing between agents potentially useful. The nature of tasks that must be performed by agents may limit how immediately information can be utilized, requiring memory and independent action. A related pressure is limitation on the lexicon size available to the agents for communication. This could require generalization and furthermore may be a natural consequence of coding constraints on noisy information transmission (see Nowak et al., 1999, for a complete discussion). Another important pressure might be temporal: environment dynamics might require speed or brevity in communication.

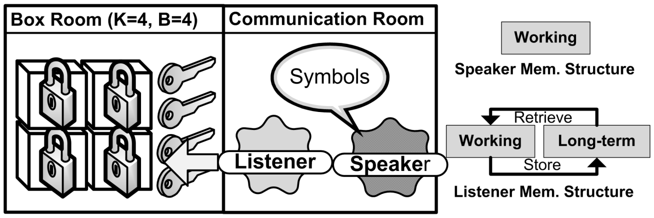


Figure 1: Treasure box domain.

Identifying structural constraints on agents is a second major requirement for this approach. These constraints may be independent of learning mechanisms and describe computational and physical capabilities of an agent. Our interests initially are in cognitive and perceptual constraints, such as limited attention and short-term memory. In the experiments below we adopt highly idealized versions of such constraints, but we always define computationally complete agents that can condition their control of internal and external processes on an internal state that combines memory and perception.

One concern about this approach is the prospect that pressures in the real world and human cognitive capabilities are so complex that our proposed analysis is impossible. However, this is an empirical question. It could very well be that careful investigation will yield simple features or ones that can be idealized while retaining their important aspects. It could very well be that careful investigation will yield simple features or ones which could be idealized while keeping their important aspects. It may also be possible to separate and explain specific language properties on a large scale.

Example: Treasure Box Domain

To demonstrate this approach to understanding language emergence we designed a set of experiments in which particular kinds of communication may emerge as optimal (or approximately optimal) behavior in a simple domain populated by two computationally limited agents. We describe next the structure of this domain and then discuss why it is of potential interest for our purposes—why we expect interesting linguistic systems to emerge.

Environment and agent structure

Figure 1 shows the Treasure Box domain. There are two agents, SPEAKER and LISTENER, who share the goal of opening a locked treasure box. These agents are in an environment containing two rooms: a first room, *communication room*, in which LISTENER can hear symbols uttered by SPEAKER and a second room, *box room*, in which there are *B* different boxes and *K* keys. At any one time, only one particular box contains treasure and can only be opened by one particular key. To solve this problem, LISTENER must go into *box room* and choose the correct box and key. However, LISTENER knows neither which box contains treasure nor which key opens it. The second agent, SPEAKER, knows the correct box and key, but cannot leave the *communication room* and therefore cannot open the box itself. Instead, SPEAKER can communicate with LISTENER by uttering symbols from a lexicon of size S which LISTENER observes while in *communication room*.

When SPEAKER utters a symbol it is placed into LISTENER's immediate perception: a buffer holding a single symbol (working memory). In addition to the working memory store, LISTENER has a second memory location to hold a single symbol (long-term memory), the value of which cannot be observed without retrieving it. LISTENER can move a symbol from the working memory store into long-term memory and vice-versa (memory encoding and retrieval), but can only observe the symbol in working memory. The agent does, however, know whether long-term memory contains information. SPEAKER remembers the last symbol uttered in an observable working memory.

Speaker. This agent observes: (1) The box containing treasure; (2) the key which opens that box; and (3) last symbol it uttered. It can act by either (1) waiting or; (2) uttering a single symbol out of a limited set of size S.

Listener. This agent observes: (1) the room it is in; (2) whether it holds a key; (3) whether it holds a box; (4) whether its long-term memory contains information; and (5) the contents of its working memory. It can act by (1) moving to the box room; (2) encoding a symbol from working memory into long-term memory; (3) retrieving a symbol from long-term memory into working memory (4) picking up a specific key; or (5) picking up a specific box.

Dynamics. The domain is structured as an episodic task where each episode ends when LISTENER picks up both a box and a key (at which point the key is automatically used to open the box). If the key is correct and the box and the box contains treasure then *both* agents will receive a positive reward (of $+1$); otherwise no reward is received and a new episode begins. At the beginning of an episode the box containing treasure and the key that opens it are chosen randomly, LISTENER is returned to *communication room* holding neither key nor box, and both agents' memories are cleared.

Learning algorithm. Although the specifics of the learning mechanism are not the focus, we needed a method for discovering good agent behavior. Both agents use the ε-greedy Sarsa(λ) algorithm (Sutton & Barto, 1998). This algorithm learns by estimating state-action values $Q(s,a)$ that represent the best expected discounted sum of rewards over an episode that can be gained by following action a from state s and then the best policy thereafter (we initialize the Q values to 0). At each step actions are chosen greedily based on the current Q function except with a probability of ε when a random action is chosen instead (yielding exploration). We use a low exploration rate of $\varepsilon = 0.01$ across our experiments. After action a_t in state s_t at time t, the algorithm updates the Q value for all state-action pairs (s,a) according to their eligibility $e_t(s,a)$ as follows earlier actions by

$$Q_{t+1}(s,a) \leftarrow Q_t(s,a) + \alpha\delta_t e_t(s,a),\ \forall s \in S, \forall a \in A$$

where before the update $e_t(s_t,a_t)$ is set to 1.0 and the eligibility for every other state-action is decreased by a multiplicative factor of γ, λ (we used $\lambda = 0.8$ for all of our experiments); the more recently a state-action pair is visited the higher its eligibility and the more credit or blame it gets for the temporal difference error $\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$ which is the the current estimated value of the resulting (s_{t+1}, a_{t+1}) plus the reward r_t immediately gained minus the predicted value of the pair (s_t, a_t). The discount factor γ describes how much less future reward is valued compared to immediate reward; we used $\gamma = 0.8$ for all our experiments. The step-size parameter α controls how fast the algorithm incorporates new experience, we use $\alpha = 0.03$ in all of our experiments.

Why this domain is of potential linguistic interest

Without any communication the best LISTENER can do is to open an arbitrary box with an arbitrary key. Given KB possible box-key combinations the probability of success at each episode is $\frac{1}{KB}$. To improve beyond this, a communicative policy is required wherein SPEAKER informs LISTENER of the correct box and/or key in some way.

Different environmental pressures and agent constraints make different behaviors optimal. For example, we can explore how varying the size of the available lexicon alters behavior. If there are enough symbols ($S \geq KB$), then a single symbol suffices to describe each box-key combination. If there are at least $K + B$ but fewer than KB symbols, then two symbols are required but each box and each key could be given a unique symbol removing the need for symbol order. Finally, with $S = max(K,B)$ the meaning of symbols will have to be shared between boxes and keys, so order may be important. In all cases these interpretation of the symbols must be learned by both agents.

We can explore the effects of changing other constraints as well, such as agents' memory or environment structure. For example, if LISTENER can store two symbols in working memory, then consistent symbol order may not matter. If the environment is no longer divided into two rooms (so communication and box opening can occur simultaneously) symbol order might still matter, but the LISTENER may not need to encode anything into long-term memory, instead acting based on the contents of its working memory at every step—in effect becoming a situated instruction-taker.

Linguistic Properties of Emergent Policies

We conducted three sets of experiments (eight individual experiments) to demonstrate how environmental pressures and agent constraints jointly effect communication properties; the experiment structure and results are summarized in Table 1. In all experiments the number of boxes and keys is equal $K = B = 4$. The first set is the domain originally described with two separate rooms where LISTENER has a working memory of one symbol and a long-term memory of one symbol. The second set modifies the agent constraints by giving the LISTENER two symbols in working memory (no long-term memory). The third set changes the environmental pressures by removing the room separator.

Table 1: Summary of three sets of experiments and policies learned. See text for detailed description.

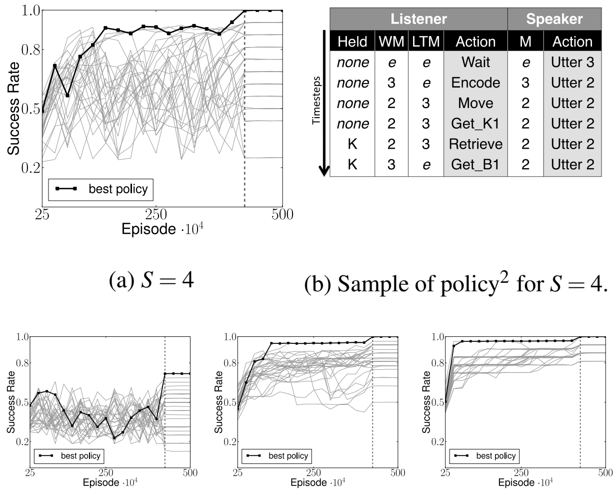
ENVIRONMENT	AGENT MEMORY	LEXICON SIZE (S)	PROPERTIES OF EMERGENT LINGUISTIC SYSTEM
Two Rooms	one symbol working memory + one symbol long-term memory	3	Association and systematic order, where in addition single symbols uttered in isolation denote specific box-key combinations. Can only achieve 75% success.
		4	Association and systematic symbol order. SPEAKER first describes the box, then the key (see Figure 2b).
		8	Highly context-dependent and idiosyncratic symbol meanings. For example *key 2* is represented by *symbol 4* if uttered before box, but *symbol 5* after.
		16	Each symbol denotes a box-key combination. For example symbol 5 means *key 1* and *box 1*.
Two rooms	two symbol working memory (no long-term memory)	3	Similar to case with 3 symbols above.
		4	Complex lexical forms. Describes entire box-key combination with two symbols which can be observed simultaneously by LISTENER effectively creating a 2-symbol length word (see Figure 3b).
One room	one symbol working memory + one symbol long-term memory	3	Symbols act as direct orders to LISTENER, but otherwise policy is similar to the cases of 3 symbols above.
		4	Association and symbol order, but no storing or retrieving from long-term memory is necessary because LISTENER can act immediately upon hearing a symbol (see Figure 4b).

Experiment set 1: Exploring constraints on the lexicon. We explore four different lexicon sizes: $S = 16$, $S = 8$, $S = 4$, and $S = 3$. Figure 2 shows 30 independent learning trajectories for each value of S. The high variance is due to the nature of the learning algorithm which may not converge for both agents every trial (or may get stuck on a less-than-optimal policy)—but what we are interested in are the best policies learned (because the mechanism used can be improved significantly beyond our initial implementation of Sarsa(λ) with fixed parameters across all experiments).

The first four rows of Table 1 summarize the results. Here we will discuss the resulting policies in more detail. For 16 available symbols, as expected, a different symbol is associated with each box-key combination and the agents arrive at perfect performance. With eight symbols, again the best performing policies use two-symbol utterances for each box-key combination, but not always in the same order (i.e. for some combinations keys are uttered first and in other boxes are uttered first). For the case of four symbols, the best performing policies communicate box and key in a particular order, with each symbol able to refer to either box or key (see Figure 2b). Of particular interest is that the the agents settle on a consistent order across box-key combinations, but this order might be different over seperate experiments: the linear position is necessary but the specific order is not. Finally, for the case of only three symbols the agents again learn a policy where linear symbol order matters. Curiously, this alone should only afford success in 56% of combinations; some policies however achieved 75% success. The policy succeeds in the additional box-key combinations by associating each with a single symbol uttered in isolation. That is, with limitations in symbol size utterance length becomes informative in addition to positional information.

As we can see, this method of systematically altering only a single constraint (lexicon size) yields broad variation in linguistic properties even in this extremely simple domain, including the denotation of symbols and the use of order information. The case of three and four symbols suggests that limited memory (paired with environmental pressures) leads to the systematic use of symbol order in optimal performance, especially when the lexicon size is limited.

Experiment set 2: Modified agent constraints. Here our aim is to explore further what specific constraints led to the systematic use of order in Experiment 1. We alter the constraints on the agents by allowing the LISTENER two symbols in working memory instead of one (and no long-term memory). All the other dynamics of the Treasure Box Domain are kept constant. The actions of *store* and *retrieve* have new

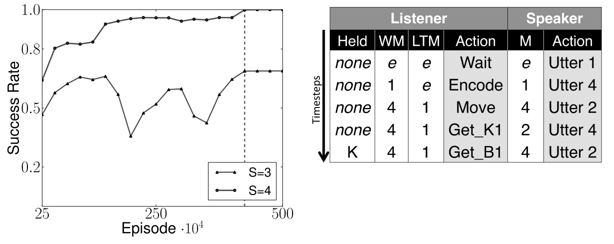


Listener				Speaker	
Held	WM	LTM	Action	M	Action
none	*e*	*e*	Wait	*e*	Utter 3
none	3	*e*	Encode	3	Utter 2
none	2	3	Move	2	Utter 2
none	2	3	Get_K1	2	Utter 2
K	2	3	Retrieve	2	Utter 2
K	3	*e*	Get_B1	2	Utter 2

(a) $S = 4$ (b) Sample of policy[2] for $S = 4$.

(c) $S = 3$ (d) $S = 8$ (e) $S = 16$

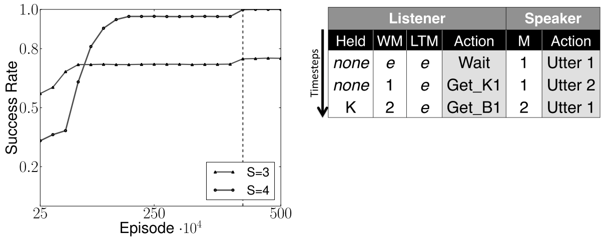
Figure 2: Experiment set 1: Exploring constraints on the lexicon. Each figure shows 30 learning curves in the Treasure Box Domain with $B = 4$ and $K = 4$. Success rate at each point is the average success rate over all episodes since previous point. Dotted line marks where learning and exploration are disabled. Best policy is highlighted and described in table 1. Figure (b) is a sample policy for $S = 4$ showing the significance of symbol order. In this case, agents have learned to associate string ”3,2” with key 1, box 1, as can be seen in the rightmost column where SPEAKER utters first symbol “3” then symbol “2”.



Listener				Speaker	
Held	WM	LTM	Action	M	Action
none	*e*	*e*	Wait	*e*	Utter 1
none	1	*e*	Encode	1	Utter 4
none	4	1	Move	4	Utter 2
none	4	1	Get_K1	2	Utter 4
K	4	1	Get_B1	4	Utter 2

(a) Best policies. (b) Sample of policy[2] for $S = 4$.

Figure 3: Experiment set 2: Modified agent constraints; LISTENER has two working memory locations. Left figure shows learning curves for best policies for $S = 3$ and $S = 4$. Right figure is a sample policy for $S = 4$ showing that the LISTENER can act according to the length-2 string in working memory: LISTENER's last two actions are box and key pickups without a retrieval in between, unlike the policy in figure 2.



Listener				Speaker	
Held	WM	LTM	Action	M	Action
none	*e*	*e*	Wait	1	Utter 1
none	1	*e*	Get_K1	1	Utter 2
K	2	*e*	Get_B1	2	Utter 1

(a) Best policies. (b) Sample of policy[2] for $S = 4$.

Figure 4: Experiment set 3: Modified environmental pressures: no room separator. Left figure shows learning curves for best policies for $S = 3$ and $S = 4$. Right figure is a policy sample for $S = 4$. The absence of a room barrier allows symbols to act as direct orders: the ”utter 1” action by SPEAKER is followed by LISTENER's ”get key 1” on the next time step.

semantics now: moving symbols between the two working memory locations. Figure 3 shows the best trial for each case in this experiment (for lexicon size of 3 and of 4). With 4 symbols in the lexicon, pairs of symbols can be used to describe each box-key combination. This is possible because unlike Experiment 1 both symbols are visible to the LISTENER (when both stored in memory) and thus there is no need for an association of order of symbol with object type (key or box). What is perhaps surprising about this result is that the more flexible agent structure in this experiment yields a simpler communication system, whereas the putatively more sophisticated linguistic system in Experiment 1 emerges as an adaptive response to the more computationally limited agent structure.

Experiment set 3: Modified environmental pressures. Here we alter the environmental constraints by removing the separator between the communication room and the box room. This modification relieves the pressure imposed by delay between communication and utilization effectively removing the need to remember information. Instead LISTENER can act immediately from SPEAKER's instructions. Figure 4 shows the best trial $S = 3$ and $S = 4$. For the case of 4 symbols, SPEAKER's utterances act as immediate instructions to LISTENER. Word order still matters, but when a particular symbol is uttered first it may correspond to a different object (box-key) than if uttered second. Furthermore, the second symbol uttered can have different meaning depending on the context. For example if LISTENER has already chosen a box, the second symbol will be associated with a key.

[2] Example policies show actions for the case key = 1 and box = 1. Each row is one time step; *e* means empty memory location. For readability, we are showing the contents of LISTENER's long-term memory and omitting current room. LISTENER does not have a “wait” action, but instead uses an action which has no effect (e.g. “pick up a key” while in the *communication room*). The SPEAKER's utterances do not impact LISTENER after it changes rooms so these actions are unimportant.

Conclusions and Looking Ahead

We have described and illustrated a novel approach to language emergence hypothesizing that specific properties of language may be understood as features of boundedly optimal policies to control problems imposed on computationally limited agents. What makes the approach distinctive is its emphasis on the shaping of linguistic systems by the joint constraints of agent and environment structure, and the emergence of such systems as the solution to the problem of how to optimally control both cognitive and physical actions in service of task goals (rather than communication goals). This means that there is no associative learning component or any other learning mechanism beyond the reinforcement learning algorithm described above. Any associations between symbols and objects or actions are arrived at not because the agents are explicitly trying to understand each other or arrive at shared symbol-meaning mappings, but rather implicitly as joint solutions to the control problem.

Our initial experiments yielded two key results. First, we have shown that even simple environments and agent architectures give rise to linguistic systems with interesting properties, including systematically structured utterances and flexible use of limited lexical resources. Second, we have shown that changes in environmental pressures or agent constraints may yield dramatic changes in optimal communication structure. Some constraints and pressures yield communication with systematic symbol order, other constraints yield policies that break the association between single symbols and single objects in the environment. The changes to environment and agent may seem small, raising the question of how a robust communication system can emerge, but in the context of the environment we explored the modifications are quite large. We expect small changes in a complex environment would not drastically alter the resulting communication systems. Furthermore, the fact that the communication system is strongly shaped by specific constraints of the cognitive architecture is also unproblematic, because we expect such constraints to be relatively stable across conspecifics. Indeed, to the extent that language is shaped by such constraints, this is good news for the cognitive scientist, because their detailed nature is likely to be more accessible that the relevant details of the shaping environments.

Our results suggest that there is promise in developing a broad systematic framework for studying language emergence by identifying mappings between pressures, constraints, and language properties independent of questions regarding the mechanisms of evolution or adaptation. Promising future avenues include investigating the emergence of compositional mechanisms like recursion, categorical features including distinctions between nouns and verbs, or more sophisticated uses of language for representation of internal mental states.

Acknowledgments: This work was supported by NSF grant IIS 0905146. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the sponsors.

References

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., et al. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, *59*, 1–26.

Chomksy, N. (2010). Some simple evo devo theses: How true might they be for language? In Y. H. Larskon R. K. Deprez V. (Ed.), *The evolution of human language: Biolinguistic perspectives.* Cambridge: Cambridge University Press.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt, Brace & World.

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489–509.

De Beule, J. (2008). The emergence of compositionality, hierarchy and recursion in peer-to-peer interactions. In *Proceedings of the 7th international conference on the evolution of language* (pp. 75–82). World Scientific Pub Co Inc.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, *97*(2), 179–210; discussion 211–25.

Gong, T., Minett, J. W., Ke, J., Holland, J. H., & Wang, W. S.-Y. (2005, July). Coevolution of lexicon and syntax from a simulation perspective. *Complexity*, *10*(6), 50–62.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569-1579.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*, 211-225.

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6889), 611–617.

Nowak, M., & Krakauer, D. (1999). The evolution of language. *PNAS*, *96*(14), 8028.

Nowak, M., Krakauer, D., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society*, *266*(1433), 2131.

Nowak, M., Plotkin, J., & Jansen, V. (2000). The evolution of syntactic communication. *Nature*, *404*(6777), 495–498.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–784.

Pinker, S., & Jackendoff, R. (2005). The Faculty of Language: What's Special About it? *Cognition*, *95*(2), 201–36.

Steels, L. (1998). Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In *Approaches to the evolution of language: Social and cognitive bases* (pp. 384–404). Edinburgh University Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* MIT Press.