LANGUAGE SCAFFOLDING AS A CONDITION FOR GROWTH IN LINGUISTIC COMPLEXITY

KIRAN LAKKARAJU[1], LES GASSER[1,2], AND SAMARTH SWARUP[1]

[1]*Computer Science Department*
[2]*Graduate School of Library and Information Science*
University of Illinois at Urbana-Champaign
{*klakkara* | *gasser* | *swarup*}*@uiuc.edu*

Over their evolutionary history, languages most likely increased in complexity from simple signals to protolanguages to complex syntactic structures. This paper investigates processes for increasing linguistic complexity while maintaining communicability across a population. We assume that higher linguistic communicability (more accurate information exchange) increases participants' effectiveness in coordination-based tasks. Interaction, needed for learning others' languages and for converging to communicability, bears a cost. There is a threshold of interaction (learning) effort beyond which (the coordination payoff of) linguistic convergence either doesn't pay or is pragmatically impossible. Our central findings, established mainly through simulation, are: 1) There is an effort-dependent "frontier of tractability" for agreement on a language that balances linguistic complexity against linguistic diversity in a population. To remain below some specific bound on collective convergence effort, either a) languages must be simpler, or b) their initial average communicability must be higher. To stay below such a pragmatic effort limit, even agents who have the ultimate capability for complex languages must not invent them from the start or they won't be able to communicate; they must start simple and grow complexity in a staged process. 2) Such a staged approach to increasing complexity, in which agents initially converge on simple languages and then use these to "scaffold" greater complexity, can outperform initially-complex languages in terms of overall effort to convergence. This performance gain improves with more complex final languages.

1. Introduction

Language evolution studies generally assume that the developmental trajectory for human languages followed stages from simple signaling systems to holistic protolanguages to simple compositional languages, and finally to the lexically and syntactically complex languages known today. If languages indeed grew from the simple to the complex, several questions need answering; two of these are:

- Could complex languages ever emerge early? Why or why not?
- Local, individual innovations that increase linguistic complexity also create linguistic diversity and, at least temporarily, reduce communicability. How can a population maintain the communicability of its language while accommodating the diversity of innovation?

While inspired by the enduring issues of human language evolution, we are primarily interested a *design stance*: evolving artificial languages for artificial agents. We need to discover *general* principles of language emergence that also cover automated agents with different sensorimotor, cognitive, and/or interactional possibilities from humans, their evolutionary predecessors, or animals. We believe, in fact, that language evolution is a *model problem* for issues that arise in many kinds of distributed semantic systems, including Web semantics, resource description-discovery (metadata) systems, cartographic systems, and biological systems. One case in point is the intentional creation and ongoing revision of XML-based semantic web languages. These can vary in complexity (number of terms, syntactic categories, etc.), and they exhibit frequency-dependent "network effects": any single language in the space has little value until a large population of agents can apply and interpret it. In this situation also, the two questions above are important: communities must converge on shared languages quickly, and ongoing linguistic innovations should only minimally disrupt the use of the language.

1.1. *Assumptions*

We are interested in artificial agents that operate continuously over long periods of time in complex worlds, performing tasks that require coordination. The value of (reward from) successful coordination drives information exchange, which in turn drives agents to create and share languages. While rewards actually come from doing things with shared information, we can usefully attribute at least part of the reward to the language itself. Thus a language that allows agents to exchange more critical information or to coordinate better has a higher value.

We assume that agents need to talk to each other about conditions and events in their worlds, and this talk is valuable in the sense above. The ability to *describe* and *distinguish* objects and actions are the fundamental kinds of information needed for coordination and increased fitness.

We consider task complexity to be information-theoretic. That is, tasks differ in complexity on the basis of how many different objects, situations, and actions they involve, and how much information is needed to reliably distinguish these objects, situations, and actions. This becomes important later when we discuss how to measure complexity of language. The ability to handle greater task diversity and task complexity increases agents' fitness; greater linguistic complexity helps enables this (as greater cognitive and motor complexity, etc. also would).

Since tasks of interest here require successful communication, and since what needs to be communicated for unsophisticated tasks is different ("simpler") than what needs to be communicated for complex tasks, agent communication languages have to vary with task complexity. For agents to become competent at more complex tasks, they need more complex languages. This means that languages have to change in complexity over time.

2. The complexity-diversity-effort frontier

Since collective activity is ongoing and must remain so while complexity grows, we have a difficult problem: *how do agents change their languages from simple to complex while maintaining communicability*? Language variation must originate at the individual level (Croft, 2001). If this is so, then as an agent originates a change from a fully communicative language, the agent will become less communicative with others, thus less effective in coordinated tasks. For language to grow in complexity this means there is a trajectory through which agents must somehow innovate (increasing complexity and decreasing communicability), then then build up communicability again by learning and propagating the innovations. This disruptive shift characterizes each increase in complexity.

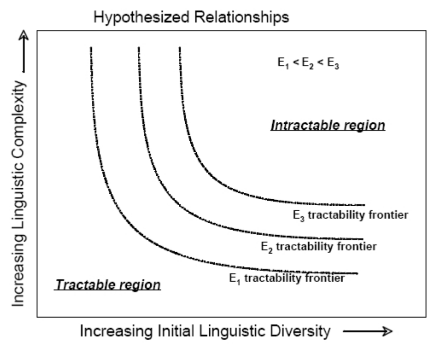

Figure 1. Conjectured tractability frontiers.

Computational tractability is an issue for this complexity growth. We hypothesize that given any set of agents with a fixed cognitive structure and a set of tasks (need for language), there exists a *frontier of tractability* for convergence to a common language. Informally, for a set of languages L of a given complexity C, greater initial diversity in the subset l of L spoken in the population will imply greater learning effort (e.g. time) to converge the population to full communicability. Similarly, for a given degree of initial linguistic diversity D, higher linguistic complexity implies greater effort to converge the population to full communicability.

Let us limit the available convergence time (i.e., effort to converge) to some amount E and plot $c = f(d, E)$, where f means "given a set of agents whose set of languages exhibits diversity d, let $f(d, E)$ equal the maximum linguistic complexity for which the population will converge within E units".

Then we will see a curve with the following property: Any complexity-diversity point "under" the curve limited by E (i.e. where for any point (c, d), $c < f(d, E)$) will converge in time bounded by E while any point "above" the curve (i.e. $c > f(d, E)$) will take longer than time E. (See Figure 1.)

E establishes a tractability frontier of complexity and diversity. *Higher* linguistic complexity *lowers* the degree of diversity the population can sustain and still converge within E. As a result, for languages that are higher in complexity, agents must make fewer, smaller innovations (introduce less diversity) if they are to converge within E.

Similarly, for a population to exhibit greater linguistic diversity and still have the possibility of converging tractably, its linguistic *complexity* must be lower. If a population is going to be highly innovative linguistically, introducing great diversity, then its language must be simple enough that the effort to converge from more widely varying linguistic "starting points" remains below E.

Throughout this discussion we focus on languages as lexical matrices. A study of convergence frontiers for structured, compositional languages (languages with a grammar) is left for future work.

3. Implementation and experiments

We demonstrate the existence of tractability frontiers through an experiment. Each agent represents its language as a *Form-Meaning Association Matrix*, which is a likelihood matrix that explicitly stores the joint likelihood of the forms and meanings. Forms are symbols in the language and meanings are concepts that can be talked about. For the present, we assume the simplest possible setup: the number of forms and meanings is equal, and the set of forms and meanings is shared among all the agents, so they are only tasked with achieving consensus on the associations between forms and meanings.

The language game proceeds through random interactions between agents. We assume a "full information" scenario, where agents provide form-meaning pairs to hearers. A speaker generates a form for a given meaning, j, by finding the element in column j of its form-meaning matrix, that has maximum value. This is a maximum likelihood rule for language production.

If σ_{ij} is the current value of the hearer's form-meaning matrix for the given symbol-meaning pair, it gets updated as follows: $\sigma_{ij} = \eta \cdot \sigma_{ij} + (1 - \eta)$. Additionally, all the values in row i are updated as $\sigma_{ic} = \eta \cdot \sigma_{ic}\ \forall c \neq j$, and all values in column j are updated in the same way, $\sigma_{rj} = \eta \cdot \sigma_{rj}\ \forall r \neq i$. This "lateral inhibition" is meant to discourage synonymy and polysemy (Vogt & Coumans, 2003).

3.1. *Measuring linguistic diversity*

In order to understand the limits of this process, it is necessary to understand how much diversity can be introduced in a population such that the population can still return to (or maintain an adequate degree of) communicability to be successful in the ongoing tasks they face.

There are several principled ways to measure linguistic diversity. Greenberg's index (Greenberg, 1956) measures diversity as the probability that a pair of randomly selected individuals from the population do not speak the same language.

$$A = 1 - \sum_i p_i^2, \tag{1}$$

where p_i is the probability of encountering a speaker of language i. Greenberg also suggests modifying this formula to take into account the similarity between languages, thus,

$$B = 1 - \sum_{ij} p_i p_j r_{ij}, \tag{2}$$

where r_{ij} is a measure of the overlap between languages i and j. A and B are both measuring communicability (or rather, the lack of it) in the population. We say a population is *converged* if the communicability is 1, i.e. diversity is 0.

Another measure, more popular in genetics, is known as the Jensen-Shannon diversity (see, e.g., Grosse et al., 2002), given by,

$$J = H(\lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_n P_n) - \sum_i \lambda_i H(P_i), \tag{3}$$

where $\sum_i \lambda_i = 1$, and the P_i are the probability distributions describing the languages (form-meaning associations). H is the Shannon entropy function. Since languages for our agents are defined as the joint likelihood matrices for forms and meanings, J measures the diversity in the corresponding probability distributions which are obtained by normalizing the form-meaning matrix. When all distributions are identical, $J = 0$.

The difference between Greenberg's index and Jensen-Shannon diversity is analogous to the difference between phenotype and genotype in biology. J is a measure based on the underlying probability distribution, and A and B are more "behavioral" measures as they directly evaluate communicability. When $J = 0$, A and B are also 0, and J attains its maximal value, A and B equal 1. However, it is possible to have perfect communicability even if the underlying distributions are not identical, since communicability depends on the maximum likelihood interpretation.

3.2. *Generating diversity*

To evaluate the tractability frontier we need to *create* a population with a specified diversity, not *measure* the diversity of a given linguistic population. To do this we initialize the agents with identity matrices for their form-meaning mappings. Then we *devolve* this perfectly converged state by adding a uniform random variable, drawn from a range $[0, \epsilon]$, to each value in the matrix. It turns out that the noise level, ϵ, is very strongly correlated with Greenberg's index and the Jensen-Shannon diversity. In other words, by increasing ϵ, we can smoothly and (nearly) linearly increase the diversity of the population according to these two measures. We have confirmed this fact through careful simulation (not presented here for lack of space).

3.3. *Linguistic Complexity*

Complexity is determined by both form and meaning complexity. McWhorter has defined four criteria for the evaluation of the complexity of a language (McWhorter, 2001), based on phonology, syntax, grammaticalization, and morphology. However, only his grammaticalization criterion makes reference to meanings. It says that a language is more complex if it makes finer semantic and pragmatic distinctions.

The language of an agent also reflects its cognitive capabilities, and an agent capable of making greater cognitive distinctions will have a more complex language simply by virtue of being able to express more meanings. This is an information-theoretic notion of complexity, as discussed earlier, and should be included in a measure of linguistic complexity. This is understandably hard to do for natural languages, but is the criterion we use in our simulations because artificial agents, in particular, can differ widely in their cognitive capabilities and characterizing this distinction is essential in a discussion of language evolution.

3.4. *Experimental result*

We measure effort as number of iterations required to converge. We initialize a population of ten agents with varying levels of diversity as described above. We also vary the complexity of the language by varying the number of meanings. Then we run the language game for each initial condition and evaluate the number of iterations necessary to converge to a communicability level greater than 0.9. This gives us a three-dimensional graph, shown in two dimensions in Figure 2, with time to convergence color-coded. We see a clear emergence of frontiers, demarcated by regions of different colors, confirming our hypothesis from Fig. 1.

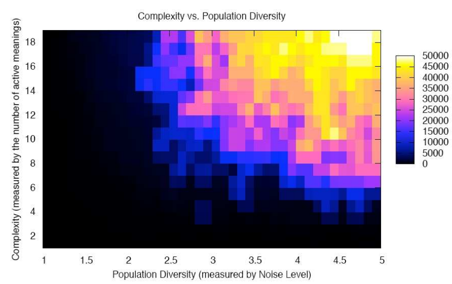

Figure 2. Time to convergence vs. complexity and diversity.

4. Scaffolding and staged learning

“Scaffolding” is one means of overcoming the diversity/complexity frontier established by E. Scaffolding is a general human learning strategy, and its existence and efficacy has been reported for language learning both in the psychological literature (Iverson & Goldin-Meadow, 2005) and in simulation work (Elman, 1993).

Lee, Meng, and Chao (2007) provide a model of “staged learning” that captures the idea of scaffolding. Agents a) constrain choices, b) act within those constraints until c) no novelty appears, then d) lift some constraints, and repeat. Constraints temporarily reduce the agents' decision space. When quiescence occurs at one stage, strategically-chosen constraints are lifted. (Thus staged learning is order-dependent and there are likely more and less effective developmental trajectories.) Learning commences again in an extended decision space, now biased by the structures and generalities learned in prior stages. We created such a staged version of our experiments as follows. We choose a maximum number of meanings, n, that the population has to converge upon. However, the agents do not consider all of these meanings initially.

They start at Stage 1. The number of *active* meanings (= "used in language games") is a function of the stage number. The *complexity step size* δ represents how many new meanings to make active per stage. Thus the number of meanings active at Stage i is δi. If the system is in Stage 4 and $\delta = 4$ there are 16 active meanings. Each agent is initialized with a $m \times n$ lexical matrix. However at each stage i, an agent only sees part of its full lexical matrix, of size $i\delta \times i\delta$. As the stages progress more of the agents' lexical matrix is revealed, as illustrated in figure 3. The system changes stages based on the communicability of the population. Let θ be the *stage transition communicability threshold*. When the population has communicability $\geq \theta$ in stage i, it has converged to within θ on $i\delta \times i\delta$ forms and meanings. It then moves to the next stage and uncovers new meanings for each agent[a]. At this transition point, $(i-1)\delta$ meanings have already been converged upon (to within θ), and δ meanings are new. These earlier convergence decisions bias agents' learning choices for the new, larger matrix. This is scaffolding. To confirm the value of staging, we repeated the earlier experiment with staging added, evaluating the new tractability frontier for varying complexity and diversity levels. Note the axes of this plot go much farther than the axes in Figure 2. In

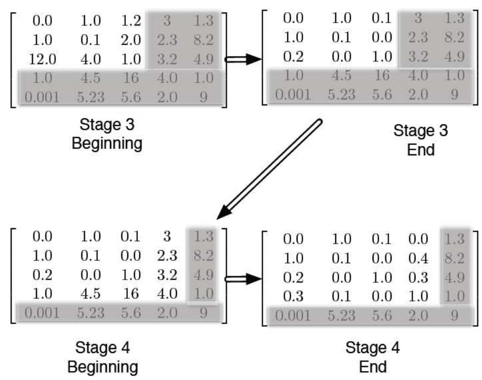

Figure 3. Moving from Stage 3 to Stage 4 uncovers a row and column of the matrix. The grey areas are hidden to the agent until it reaches that stage. $\delta = 1$

[a]Collective ordering of meanings is an issue, with several possible efficient approaches, e.g. common environment structure. We leave to future work a more detailed model exploring this topic.

fact we began each simulation with 10 meanings and 10 forms because smaller matrices converge very quickly. Even with higher initial noise levels and number of meanings going up to 30, we see that the population converges in a fairly short amount of time. Staging has pushed out the tractability frontier greatly.

5. Conclusions

We have shown the need for scaffolding in language learning to be a fundamental requirement arising from the tradeoff between complexity and diversity. The interaction between complexity and diversity leads to the existence of a tractability frontier that prevents convergence in reasonable time if the initial diversity is too high for a given complexity of language (or vice versa). However, by learning in stages, it is possible to attain convergence even on complex languages that would otherwise be beyond the tractability frontier.

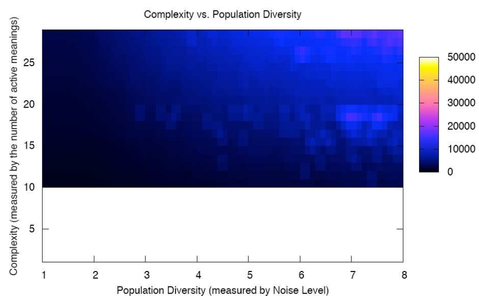

Figure 4. Tractability for staged learning.

References

Croft, W. (2001). *Explaining language change.* Longman Group United Kingdom.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71-99.

Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language*, *32*(1), 109-115.

Grosse, I., Bernaola-Galván, P., Carpena, P., Román-Roldán, R., Oliver, J., & Stanley, H. E. (2002). Analysis of symbolic sequences using the Jensen-Shannon divergence. *Phys. Rev. E*, *65*, 041905.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367-371.

Lee, M. H., Meng, Q., & Chao, F. (2007). Staged competence learning in developmental robotics. *Adaptive Behavior*, *15*, 241-255.

McWhorter, J. H. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, *5*, 125-166.

Vogt, P., & Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, *6*(1).