HOW TO DO EXPERIMENTS IN ARTIFICIAL LANGUAGE EVOLUTION AND WHY.

LUC STEELS
VUB AI Lab, Pleinlaan 2
1050 Brussels, Belgium
steels@arti.vub.ac.be, and,
SONY Computer Science Laboratory Paris

The paper discusses methodological issues for developing computer simulations, analytic models, or experiments in artificial language evolution. It examines a few examples, evaluation criteria, and conclusions that can be drawn from such efforts.

1. Introduction

The problem of the origins and evolution of language is notoriously difficult to approach in a scientific way, simply because solid data are lacking of the earliest human languages or of the neurobiological changes that enabled language. But that does not mean that scientific theorising is impossible. After all, there are many scientific fields where direct observation is not feasible, for example studies of the origins of the cosmos, and despite of this, concrete theories have been developed by analytic models, computer simulations and experiments. The same approach is possible for studying the origins of language, at least for certain aspects of this question.

In what follows, a communication system is said to be ‘natural language like’ if it has features such as: compositionality, marking of predicate argument structure in terms of abstract semantic roles and cases, use of perspective in the conceptualisation and marking of perspective, use of hierarchy and recursion, for example for grouping lexical items that share semantic functions (like the words in a noun phrase), use of pronouns or other elliptic expressions for reference to entities already introduced in earlier discourse, conceptualisation of events and marking in terms of Tense-Aspect-Mood systems, marking of information structure through syntax (e.g. a topic-comment distinction), etc.

The work considered here assumes that the needs for a complex communication system with such features are there and that at least the basic neurobiological machinery to configure a language faculty are there as well, but then asks *how* a complex, natural language like communication system might develop, specifically: (i) What kind of cognitive mechanisms individuals need to develop and sustain such a system, (ii) what factors make these mechanisms relevant for communication, and (iii) by what processes the mechanisms get configured into a language faculty. It is for this type of investigation that computer simulations and experiments in artificial language evolution are appropriate, particularly if one seeks a theoretical explanation which constrains how language evolution *might* have happened.

For over a decade now, our team has been doing computer simulations and experiments in artificial language evolution to try and explain the origins of such natural-language like features, starting from agent-based models of a spatial language game (Steels, 1995), and then branching into experiments with robotic agents able to self-organize communication systems grounded in reality through their sensori-motor apparatus (Steels, Kaplan, McIntyre, & Looveren, 2002), (Steels, 2004). Other representative work is found in collections by (Briscoe, 2002), (Cangelosi & D.Parisi, 2003), (Minett & Wang, 2005), a.o. These collections also contain various attempts to develop analytic models for aspects of language evolution.

Although those engaging in these kinds of studies feel that there is steady progress with very profound results, the impact on other disciplines interested in the origins and evolution of language has so far been limited. Reactions vary from fascination and incomprehension to scepticism or downright rejection. These reactions are partly due to a lack of explanation from those of us using these approaches. It is perhaps not clear how the methodology works and why it is relevant. Moreover the criticisms are to some extent justified because the model assumptions are not always very clear or downright unrealistic, and often conclusions are drawn which are not warranted based on the models that have been proposed. This paper is intended to clarify methodological issues and sharpen the criteria for their sound application. I discuss first computer simulations, then analytic models and then experiments in artificial language evolution.

2. Computer Simulations

Three steps are involved in setting up computer simulations: (1) The researcher hypothesises that a certain set of cognitive mechanisms and external factors are necessary to see the emergence of a specific feature of language. (2) The mechanisms are operationalised in terms of computational processes, and (simulated) 'agents' are endowed with these processes, (3) A scenario of agent interaction is designed, possibly embedded in some simulation of the world. The scenario and the virtual world capture critical properties of the external factors as they pose specific communicative challenges. (4) Systematic computer simulations are performed, demonstrating that the feature of interest indeed emerges when agents endowed with these mechanisms start to interact with each other. Ideally a comparison is made between simulations where a mechanism or factor is included and others where it is not, in order to prove that the mechanisms or factors are not only sufficient but also necessary. This still does not prove anything about human language evolution because there may be multiple mechanisms to handle the same communicative challenges, but at least it shows a possible evolutionary pathway.

Here is one example of this approach: the Naming Game (Steels, 1995). Every human language features proper names for individual objects, and this must have been an obvious first use of language, for example to call or designate members of the group. A crucial question is then: How can a population converge on a consistent set of names for a particular set of objects, without a prior system, a central authority, or telepathy (one individual having access to the internal brain state of another one). The Naming Game studies this question by framing interactions in terms of language games. The speaker uses a name to identify some topic in the context, and the hearer guesses the topic based on the name. The game is a success if the hearer was able to identify the same topic as chosen by the speaker.

It is now known that agents can use a wide variety of strategies to play the Naming Game, each implying particular cognitive mechanisms. For example, computer simulations (as shown in figure 1) have shown that using an associate memory of object-name pairs with weights and lateral inhibition is a good strategy. For those unfamiliar with computer simulation, it is perhaps important to stress that such simulation results do not depend on a specific computer implementation nor on the programming language used, nor even on the fact that a computer is used. The simulations simply show the behavior of a dynamical system. The assumption underlying this work (which is a fundamental assumption of science) is that the properties of the dynamical system constitute an explanation of the emergent phenomenon, the same way oscillations in predator-prey populations are explained by the dynamics of the Lottka-Volterra equations and depend in no way on the specific organisms involved.

For the computer simulation to have value, some conditions must be met: (1) It must be clear what language features are supposed to be emergent and what features are assumed. It is simply not possible to explain everything at once. A lot of scaffolds in terms of assumed cognitive abilities, interaction patterns or environmental constraints must be introduced. For example, the Naming Game strategies discussed earlier assume that both agents are able to individually recognise the objects they are naming, that the hearer has a way to indicate what topic he has guessed, that agents can recognise and reproduce the names used by others, and so on.

(2) There must be no hidden 'global hand', in other words effects of global properties not observable by individual agents, nor any direct causal link between a mechanism and the feature to be explained. For example, genetic models of lexicon convergence (as opposed to cultural models as discussed above) often introduce a fitness function which is calculated in terms of the similarity of an

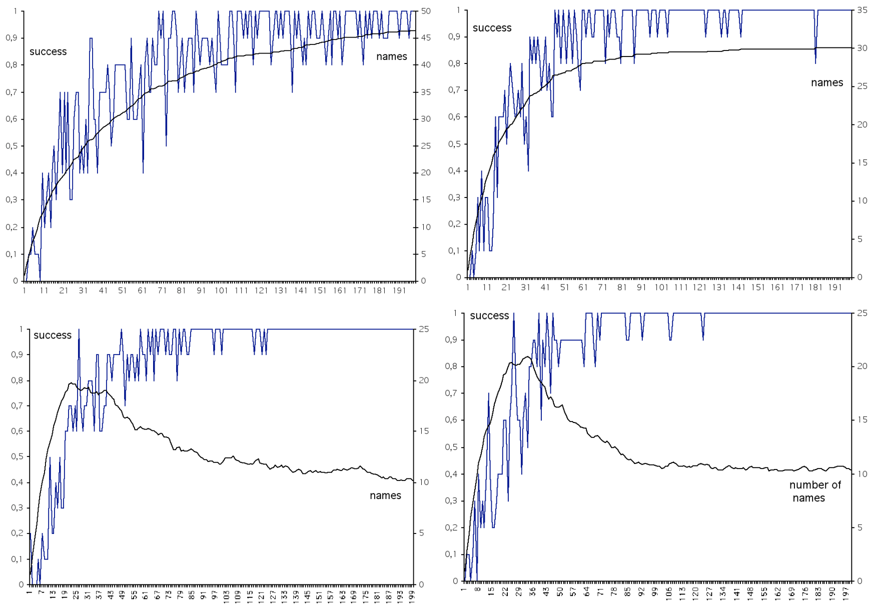

Figure 1. Effect of different strategies for playing the Naming Game. The size of the population $\mathcal{N}$ and the number of objects $\mathcal{O}$ is always equal to 10. The evolution in communicative success (left y-axis) and average inventory size (right y-axis) is shown for 2000 games (x-axis). Top left shows a strategy where agents simply adopt the word used by others. After a while everybody knows all words and hence there is complete communicative success but the inventory is large (45 words). Top right shows a strategy where success translates to enforcement (weight increase) of the word used. Success is reached more quickly and the inventory size goes down (30 words). Bottom left adds lateral inhibition (decrease of weight of competitors) and bottom right adds damping (weight decrease in case of failure). The last strategy leads to an optimal inventory (10 words) and fastest convergence, while tolerating homonymy.

agent's lexicon with that of others in the population. The computation of this fitness function requires a global view which none of the agents can have. The same sort of models also try to explain convergence by setting up a selection process that is based on greater fitness, but this fitness is calculated in terms of similarity of lexicons, in other words on how well the lexicon of the agent converges to that of the group. So there is an undesirable direct causal link between a (global) mechanism and the feature being explained.

(3) It is crucial to consider not only configurations that 'work' but also those that do not work or work less well, both to understand the causal role of each specific component integrated in the language faculty of the individual agents, and the role of parameter choices for the different mechanisms (as shown in figure 1) or the environmental factors. All this is standard scientific practice (Platt, 1964) and can be applied easily here.

3. Analytic Models

Computer simulations are an effective way to test claims about the sufficiency and necessity of certain cognitive mechanisms or how communicative challenges impact the evolution of a language, and are very valuable because it is notoriously difficult (even for computer scientists) to understand how specific computational mechanisms affect the outcome of observed (collective) behavior. But computer simulations have a major limitation. They cannot predict the general long-term behavior of a system. This is where analytic models come in. They aggregate the state of individual agents or agent behaviors by postulating global quantities with which a series of master equations is formulated. Then the standard mathematical techniques for solving these equations can be used to predict the global time course of the system. Of particular relevance is the search for scaling laws, which capture how increase in certain system parameters (for example the number of agents in the population, the number of objects they have to name, etc.) impact other system properties (such as time to reach convergence, size of the lexicon, etc.). Normally, the global quantities used in analytic models are measured by empirical observation, but, if data is missing, as in the case of language evolution, the approach can be applied to the outcome of computer simulations.

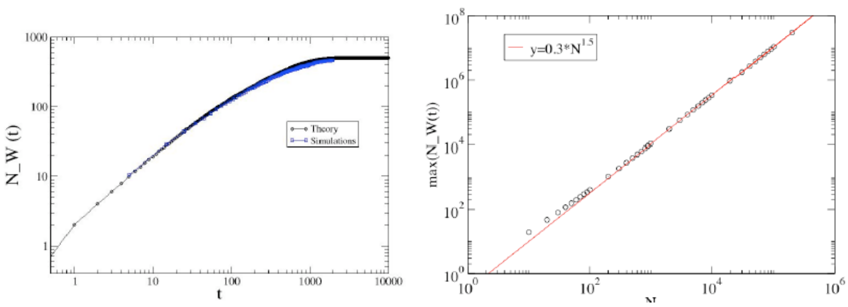

Figure 2. Very close fit between a simulation and an analytic model of the Naming Game (left). Power law behavior of the Naming Game is shown in log-log plot (right). The maximum number of words (y-axis) has a power relation with population size (x-axis) with exponent 1.5. It is not only observed in computer simulations but also predicted by the analytic model.

A recent example of this approach for the Naming Game in very large populations is discussed in (Baronchelli, Felici, Caglioti, Loreto, & Steels, 2005). It focuses only on naming one object and uses global quantities like the number of agents N_a, the total number of words at time t $N_w(t)$, the number of different words $N_d(t)$, the success rate $S(t)$, and the overlap function $O(t)$ which monitors lexical coherence in the system. It is possible to analytically predict the behavior of these global quantities from master equations using a mean field approach (figure 2(left)) and to identify power laws, such as the one shown in figure 2 (right), and prove why they have these exponents. In this type of investigation, the role of the computer is restricted to calculating the graphs that display the mathematical functions derived from the equations. These are not computer simulations. Models of agents have completely disappeared.

There are some criteria that analytic models must meet in order to be relevant: (1) The models must in one way or another relate to data, ideally from empirical sources but otherwise at least from computer simulations. Otherwise, any kind of relation can be claimed and any kind of conclusion can be drawn. Unfortunately most analytic models of language evolution that have been published so far do not meet this criterion (although the work reported above does albeit only w.r.t. simulated data).

(2) Realistic assumptions must be made about the cognitive capacities of the agents or the effects of natural or cultural selection. Human beings, as embodied autonomous agents, have strong limitations, for example, they cannot perceive the world exactly from the viewpoint of another agent and so equal perception is excluded, direct meaning-transfer is not possible, no agent can have a global overview of the language in the total population, grammar induction is always influenced by the available data, etc.

There are strong limitations to the analytic method, partly because the aggregate quantities and master equations must be found, which is very non-trivial, but more importantly because for a large number of non-linear dynamical systems (and language definitely falls into this category), no solution method is available or can ever be found. New techniques from statistical physics, such as network analysis offer nevertheless hope that much more is possible than so far achieved.

4. Experiments

Many empirical sciences use a third method for investigating natural systems, namely experiments. Normally, an experiment takes an existing natural system (for example a cell or a block of ice) and examines what happens when certain environmental parameters or system components are changed. An experiment therefore generates new data that would otherwise not be observable. The method is particularly appropriate to understand and prove which causal relations exist between the changed parameters and the observed system behavior. For example, between the surrounding temperature and the phase transitions of the block of ice into water and steam.

We might in principle invent experiments for language origins and evolution as well, although it is not so obvious. It is not possible to selectively turn on and off components in the brains of groups of humans and see the effect on the language that emerges in the group, or to make a group forget some aspect of their language (like the Tense-Aspect-Mood system) and see whether they evolve a new TAM system. Sometimes there are natural experiments: Brain disorders due to genetics or aging may lead to language disorders. Unusual social circumstances like rapid population change in highly multi-lingual settings may give rise to new languages or language features as in creoles. But these natural experiments are generally not sufficiently controllable for being a solid basis for doing science. Quite recently some psychologists have begun to study the emergence of communication systems in dialog by constraining normal communication or creating unusual challenges (Healey P. & Katagiri, 2002). These experiments are more controlled and yield fascinating data that are highly relevant to the question of language origins. They show for example that humans can quite quickly negotiate new communication systems and that they constantly adapt their language systems at all levels to those used by others involved in the same dialog.

However the state of the art in robotics and Artificial Intelligence makes it now possible to do non-trivial experiments with physically embodied agents (robots). Rather than selectively adding or removing components in the language faculty of humans, we do it with robots. Moreover we can control the robot's perception of the world, progressively introduce communicative challenges, control the in- and outflow of the population, the degree of noise and stochasticity in sound transmission and reception, and so on. In addition we can completely monitor both the external behavior, the emergent language system, and the internal states of the agents, even for very large populations. Such experiments in artificial language evolution have some characteristics in common with computer simulations, but they go far beyond them. Computer simulations can introduce all sorts of scaffolds and make various kinds of assumptions which can no longer be made in these experiments. For example, if we require that agents can identify objects to play the Naming Game, then we must implement the necessary perception and memory functions to achieve this - a very non-trivial task in itself. So the experiments are the most powerful and stringent way to test the realism of model assumptions.

Here is an example experiment discussed in more detail in (Steels, Loetzsch, & Bergen,). The experiment focuses on perspective reversal, a clear universal feature of human languages. A communication system with perspective reversal allows that a scene is conceptualised from different points of view (the speaker, the hearer, other participants, landmarks) and that the perspective is possibly marked explicitly, as in English *your left* versus *my left*. The perspective reversal experiment uses two autonomous AIBO robots that move around in search for a ball and, if they have found one, play a description game, describing to each other the movement of the ball, such as 'the ball was far away to my right and then rolled to your left' (see figure 3). The population starts without any perceptual categories (like left/right or close/far) and without any lexicon, but has to evolve sufficiently shared ontologies and lexicons to be successful in the game.

The perspective reversal experiment examines three issues: (1) Why is perspective reversal needed. It turns out that agents can develop an adequate system

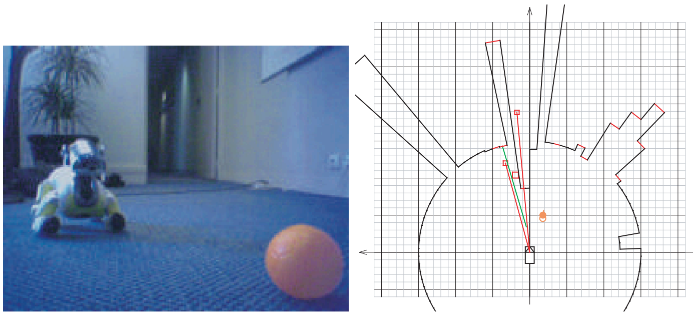

Figure 3. AIBO robot used in perspective reversal experiment (right). The dynamic world model of the robot as it is tracking the ball, obstacles and other robots. The description game is based on such world models.

if they (unrealistically) share exactly the same perception (figure 4a) but as soon as they see the world from their own perspective - which is always the case in embodied agents - their communication system collapses (figure 4b). (2) Then agents are given the ability to perform egocentric perspective transformation, which means that they can geometrically transform their perception of the world to see the scene from the viewpoint of the other agent, and they use that in conceptualisation. Communicative success goes up again (figure 4c). (3) Next agents also mark perspective which means that information flows from the egocentric perspective transformation component to the lexical component. We see that cognitive effort goes down (figure 4d). This experiment therefore demonstrates why we see perspective reversal and marking in human language: it increases communicative success in the case of embodiment and decreases cognitive effort.

As before we list a few criteria that such experiments need to take care of: (1) We first of all get the same caution as with computer simulations: It must be clear what feature is supposed to emerge and experiments must be done to compare configurations with and without cognitive components held responsable for their emergence and in different environmental circumstances (as in the experiment in figure 4).

(2) There should obviously be no global hand nor any direct causal link between mechanisms and features. It is more difficult now to make errors (compared to simulations) because certain short-cuts are no longer possible, even though scaffolds are still necessary and no problem if they do not impinge on the basic point of the experiment (for example give all robots the same perceptual system or implement the script for playing the language game).

For those not familiar with robot experiments, we need to stress that results (as those shown in figure 4) are on the one hand related to the specific embodiments (the shape, perceptual capabilities, computer processors, etc. of the robots used) however those details are not crucial, they are only instantiations of basic princi-

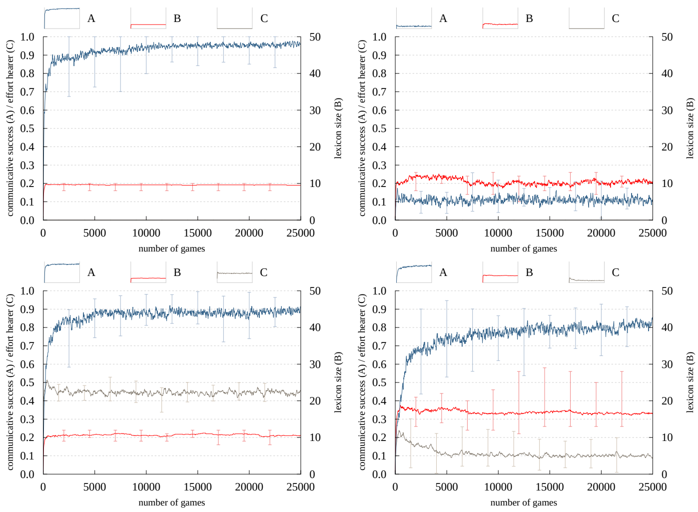

Figure 4. Experiments in perspective reversal with same and different view on scene (top), and with egocentric perspective transformation for conceptualisation (bottom left) and with marking (bottom right).

ples. Just like one can study a fruitfly to study genetic mutation rates in general. The same experiments could be carried out on other kinds of robots or even for other sensory domains or perceptually grounded categories, as long as the agents get differents views and hence different perceptions of the world so that perspective reversal becomes necessary. The specific implementation of the cognitive components is irrelevant, it is the functionality of the component that counts, and the experiment proves that these functionalities can be operationalised and that they can be put together in a way that effectively leads to an emergent communication system with this specific feature.

5. Conclusions

There is a growing number of computer simulations, analytic models, and experiments in artificial language evolution which shine new light on the age-old question of the origins of communication systems with the features of human natural languages. A large number of issues has not been tackled yet and we only have solid results so far for some of the most basic questions, such as how can a population develop a shared set of names. So this presents enormous opportunities for young researchers coming in the field. At the same time useful dialog is already possible and ongoing with the other approaches to language evolution, that emphasise the linguistic and anthopological data or constraints from neurobiology.

Acknowledgements

This research was funded and carried out at the Sony Computer Science Laboratory in Paris with additional funding from the EU FET ECAgents Project IST-1940.

References

Steels, L., Loetzsch, M., & Bergen, B. Why human languages mark perspective. In *In review.*

Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., & Steels, L. (2005). *Sharp transition towards shared vocabularies in multi-agent systems.*

Briscoe, T. (2002). *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge, UK: Cambridge University Press.

Cangelosi, A., & D.Parisi. (2003). *Simulating the evolution of language.* Berlin: Springer Verlag.

Healey P., I. U., M. Swoboda, & Katagiri, I. (2002). Graphical representation in graphical dialogue. *International Journal of Human-Computer Studies*, *57*, 375–395.

Minett, J., & Wang, W. S.-Y. (2005). *Language acquisition, change and emergence: Essays in evolutionary linguistics.* Hong Kong: City University of Hong Kong Press.

Platt, J. (1964). Strong inference. *Science*, *146*, 347–353.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*.

Steels, L. (2004). Constructivist development of grounded construction grammars. In e. W. Daelemans (Ed.), *Proceedings annual meeting of acl* (pp. 158–181). Barcelona: ACL.

Steels, L., Kaplan, F., McIntyre, A., & Looveren, J. V. (2002). Crucial factors in the origins of word meaning . In A. Wray (Ed.), *The transitions to language* (pp. 252–271). Oxford: Oxford University Press.