THE EMERGENCE OF A LEXICON BY PROTOTYPE-CATEGORISING AGENTS IN AN INFINITE WORLD

CYPRIAN LASKOWSKI

Language Evolution and Computation research unit
University of Edinburgh, Edinburgh, EH8 9LL, UK
cyp@ling.ed.ac.uk

Over the last decade, computational models and simulations have been used to explore whether words could have initially become grounded and established in the earliest stages of language evolution through a process of self-organisation in a population. In this paper, a new model of this family is produced, with two major differences from previous models: the agents' world consists of an infinite number of objects, and the agents' categories have a flexible prototype structure. These changes result in a more realistic model, but also one which is more likely to fail. Simulation results revealed that coherent lexicons still emerged, but they were sensitive to certain model conditions, including the structure of the world.

1. Introduction

Words pose an enigma for language evolution, for they are both fundamental and complex. On one hand, words constitute the basic building blocks of language, and on the surface they appear to be simply pairings of form and meaning. In fact, it makes little sense to speak of linguistic structure or its evolution without presupposing the existence of words, and their emergence is thus considered to constitute one of the earliest stages of language evolution (Jackendoff, 1999). On the other hand, words are distinguished by a set of properties that are not found together in any other animal species' signals: they are learned, arbitrary, referential and numerous. As such, words are unique to humans and cannot simply be taken as the very starting point of language evolution. Indeed, the evolutionary emergence of words is an unresolved puzzle.

Moreover, as for other aspects of language, the origins of words must be explained on at least two levels, biological and cultural. The biological level concerns questions of individual cognitive potential and linguistic preadaptations, such as a conceptual capacity. This can be partially investigated by comparing human and animal cognition, and assessing the extent to which animals can learn human words (Deacon, 1997). However, even if we pin down the necessary prerequisites, it is far from clear how the first words actually came into existence within a population of such individuals. Thus, at the cultural level, we must explain: how did hominins first invent and agree on words and their meanings?

Exploring this question empirically is difficult, since animals do not spontaneously invent words, while humans already have them. Instead, over the last decade, some researchers have tackled this problem with the help of computational models and simulations. In particular, Steels (1997) designed a simple model and showed that a coherent lexicon could emerge through a process of self-organisation. In particular, a population of individuals equipped with the necessary biological preadaptations gradually converged on a coherent lexicon by engaging in local communicative interactions about objects in a shared environment. Subsequently, new models have been implemented and have generally supported Steels' initial conclusion.

However, the validity of these findings is potentially contingent on the particular model used. In fact, previous models have tended to use simple worlds consisting of a small number of pregenerated objects, and conceptual structures which are efficient but in conflict with accepted psychological theories. It is thus possible that the models have made it unrealistically easy for a lexicon to emerge, in which case past results may have been misleading. In order to address this concern, I present here a new computational model, which adopts more complex and realistic representations (in some respects), and present simulation results.

2. A new model

The current model differs from most previous models in two important respects: the world consists of an infinite set of objects, and agents' concepts have prototype structure. However, the dynamics of the model, in terms of the agents' interaction with each other and their environment, broadly follows previous work. For a detailed description of the model and simulation results, see Laskowski (2006).

2.1. *The world*

As in previous models (Smith, 2003a), the agents' world is represented with an n-dimensional space, where each dimension can be thought to represent a perceptual feature (e.g., colour, shape, size). Objects are defined as points in this space, whose dimension values are real numbers between 0 and 1 that identify the extent to which the objects have the corresponding features.

However, although such a representation system allows for an infinite number of possible objects, previous models have typically not exploited this and have instead pregenerated a finite set of objects for each simulation, from which random contexts were then selected for each interaction (Smith, 2003a). In contrast, in this model, every agent-world interaction involves the generation of entirely new objects. As a result, even after thousands of interactions, it is extremely unlikely that an agent will ever encounter the same object twice. This is important, because in the real world, no two stimuli are exactly identical, and our concepts are flexible enough to group objects together and handle novel exemplars.

At the same time, however, the real world is not completely random, but has some structure. The world is "clumpy" (Smith, 2003b), in the sense that, *within* dimensions, some values are generally more likely than others (e.g., animals usually have even numbers of legs). Also, the world is "correlated", so that values *across* dimensions tend to correlate to some extent (e.g., things that fly tend to have feathers, and vice versa). Consequently, rather than generating an entirely random vector each time an object is needed, the objects in this model are generated pseudo-randomly in accordance with probability distribution functions defined by the model's "clumpiness" and "correlation" parameters.

2.2. *Categories*

Following Steels (1997), agents are equipped with sensory channels which detect object dimension values directly and map them onto their perceptual space (which thus have the same general n-dimensional structure as the world). Unlike objects, categories are not atomic points. In most models, however, categories are represented as well-defined regions of the perceptual space, such that objects are category members if and only if they fall inside the region. This is a simple and efficient representation system, but it is an implementation of the classical theory of concepts, which has long been considered obsolete (Murphy, 2002).

In this model, category structure is based on prototype theory (Rosch, 1978), one of the leading psychological theories of concepts, in which categories have central members and graded membership. Categories are defined as Gaussian functions over the conceptual space which assign a degree of membership (a real number between 0 and 1) to every possible object. The category's prototype is the point of maximum membership (1), and the rate at which membership decreases as one moves away one from the prototype depends on the category's sensitivity to each dimension. Formally, the category membership of an object o in a category c is given by a Gaussian function,

$$membership_c(o) = \left(\prod_{i=1}^{N} e^{-\frac{1}{2}\left(\frac{o_i - p_i}{s_i}\right)^2} \right)^{1/N} \tag{1}$$

where i identifies a dimension, with o_i being the object value, p_i the prototype value, and s_i the sensitivity. This representation is similar to that used by Belpaeme (2002), with the main difference being that in that model, the dimension sensitivities were all rigidly set to one default value.

Although this representation is relatively plausible psychologically, it makes categorisation of objects more involved. Rather than identifying the category in whose space an object falls, it is now necessary to find the category which best fits the object (i.e., the category for which the membership function yields the highest value). Moreover, a minimum threshold is defined (as a model parameter) so that an object can only be potentially considered as a member of a category if its degree of membership is above this threshold. Figure 1 shows an example of such a “candidate category” in two-dimensional space for a particular object.

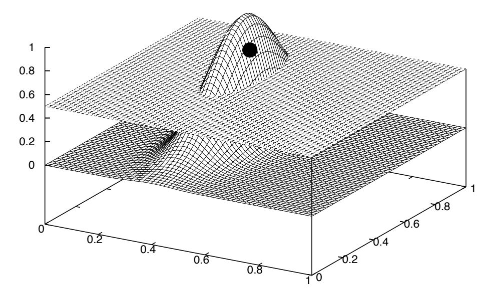

Figure 1. Category membership in 2 dimensions: $membership_c(o)$, the category membership function for an agent’s category c in a conceptual space of two dimensions, with $p_0 = 0.4$, $s_0 = 0.05$, $p_1 = 0.6$, and $s_1 = 0.1$. The plane shows the value of the minimum membership threshold, and the dot indicates the object being categorised: since the dot is above the plane, this is a candidate category for the object.

Each category is also associated with a list of words and association strengths. Words themselves are atomic tokens with no internal structure. The word with the highest association strength is the best or most “natural” word for that category, and is the word that the agent will typically use when communicating about the category. The list can also be empty, in which case the category has not been lexicalised.

2.3. *Category development*

Agents develop and adapt their category systems through interactions with the world. Each interaction takes the form of a discrimination game (Steels, 1997), in which an agent is exposed to a context of objects, attempts to find a distinct category for one of the objects (called the topic), and adapts its category system accordingly. Over many discrimination games in different environments, an agent’s category system gradually grows and adapts to the structure of the world.

Discrimination games have 3 basic possible outcomes: the creation of a new category, the splitting off of a subcategory, or the adjustment of an existing category. If the agent has no candidate categories for the topic object, then it will create a new category, whose prototype is set to the topic object, and whose initial dimension sensitivities are a function of how similar the other context objects were to the topic in the different dimensions. Otherwise, it will check whether any of its candidate categories are sufficiently discriminating as not to match any of the other context objects. If there are no such categories, then it takes the most refined candidate category (i.e., the one with the most sensitive dimensions), and creates a subcategory of it which is identical with it except for being more sensitive in the dimension in which the topic differs the most from the other context objects. If there were discriminating categories, then the topic is categorised with the one for which it has the highest membership, and this category's prototype and dimension sensitivities are adjusted slightly to fit the topic better.

2.4. *Lexical development*

The other kind of formal game that the agents engage in is the guessing game Steels (1997), which is actually built on top of the discrimination game. While discrimination games involve only one agent and do not involve any linguistic exchange, the guessing game is a communicative episode involving two agents and a shared environment. The "speaker" agent utters a word for one of the context objects (the topic), and the "hearer" agent guesses which object the speaker was referring to. The game is a success if and only if the hearer guesses correctly. Over many such games between pairs of agents from a population, the agents may gradually converge on a sufficiently similar lexicon to have generally successful communication (this is in fact the research question).

The mechanism for each guessing game is as follows. A speaker and a (different) hearer are chosen from the population at random, and a new shared context of objects is generated. The speaker chooses a topic object at random, categorises it (via a discrimination game), and utters the word in its lexicon with the highest association score for that category. If it has no word for that category, it randomly invents a new word. The hearer must find the best match between the word heard, a context object, and a word-category pair from its own lexicon. It first identifies all of its categories which have an association for the word. If there are no such categories, the game fails. Otherwise, it considers each possible category-object pair from these categories and the context objects, and determines the combination for which category membership is highest. If the resulting membership is below the minimum membership threshold, then the game is a failure. Otherwise, the hearer guesses the object from that pair. If this object is the topic, the game is a success. Otherwise, the speaker points out the topic to the hearer (non-linguistically), and the hearer performs a discrimination game on it. Upon completion of the game, both agents independently update their lexicons, increasing or decreasing specific word-category association strengths in accordance with the results of the game.

3. Simulations

Simulations were run within this model with three questions in mind. First, would agents converge on a coherent lexicon, despite the more complex world and category representations used in this model? Second, do the simulation results depend substantially on the specific world structure used? Third, assuming that agents did converge, how stable would the results be if one varied specific parameters, such as population and context size?

Each simulation consisted of a large number of guessing games in a fixed population of agents, who all began with empty category systems and no lexicons. The guessing games were analysed in sets of 100 called epochs, and the average success rate (the ratio of successful to total number of guessing games) was tracked for each epoch. The first set of simulations explored whether this model would work at all in the simplest cases, with population and context sizes of 2 in a 1-dimensional world. After 200 epochs, regardless of how clumpy the world was (dimension correlation does not apply of course in a one dimensional world), communicative success in the final epoch averaged at around 99% over 100 simulations, despite the fact that the agents ended up with large category systems.

In a 3-dimensional world, the final communicative success was still very high, despite extreme manipulations of world structure. Four kinds of world were tested: "random" (dimensions were completely uncorrelated and non-clumpy), "correlated" (highly correlated dimensions but completely non-clumpy), "clumpy" (completely uncorrelated but highly clumpy), and "structured" (highly clumpy and correlated). Final communicative success after 200 epochs was still very high for all four world structure types, ranging from 96% for the random world and 99% for the structured world. Agents ended up with around 300 categories, except for the totally random world, where they tended to have over 500 categories.

In order to explore the scalability of the results and their potential dependence on a particular world structure, further sets of simulations were conducted. In each set, one of the main model parameters was manipulated, starting with a base case of a 3-dimensional world, 2 context objects, and 2 agents. Results showed that communicative success was significantly affected by manipulations of these variables, but the extent of the impact depended on the world structure. For instance, world dimensionality had a very large impact, such that in an 8-dimensional world, final communicative success rates tended to stay below a dismal 25% in the random and clumpy worlds. However, they were still in the high 90's in the correlated and structured worlds. Manipulations of the context size had similar impacts, although less drastic. Context sizes of 64 objects still yielded approximately an 80% final success rate in the correlated and structured worlds, but context sizes of only 16 objects resulted in rates below 50% for both random and clumpy worlds. The effects of population size changes did not follow the same pattern, however. Although higher population sizes corresponded with lower communicative success rates, the final communicative success rate reached at least around 75% in all four world structures even with as many as 128 agents, and was best in the clumpy world (around 90%). Moreover, the communicative success rate curves also varied in clear ways between the four world types examined. In worlds with correlated dimensions (i.e., the correlated and structured worlds), communicative success rose very quickly (e.g., to about 75% with 128 agents), but then flattened out. In contrast, worlds with clumpy dimensions started off more slowly, but their communicative success rate curves did not flatten out as dramatically, so eventually they obtained higher success rates (at least in the clumpy world).

4. Discussion

Despite the use of a more complex model, in which agents never saw exactly the same stimuli twice and their categories had a continuous prototype structure, simulation results were generally in line with those of previous work. Under a variety of conditions, populations of agents converged onto coherent lexicons after engaging through repeated communicative episodes in shared environments. Although each agent had an independent category system and lexicon which started out empty, communication success rates managed to reach high levels, often close to 100%. These basic results, then, add support to the idea that a population of hominins equipped with certain cognitive preadaptations could have grounded and developed a large system of learned, arbitrary, referential words through a series of local interactions (Steels, 1997). More specifically, they show that the general results of previous models cannot be dismissed on the grounds that they used psychologically implausible category representations. Future models can move in both directions: they can either develop increasingly realistic category representations and test whether the basic results will continue to hold, or they can revert to simpler representations and test more complicated conditions, since it seems that the choice of representation is not too critical to the results.

However, the simulation results were sensitive to more complex conditions, as manifested by manipulations of the model's parameters. As in previous models, communicative success dropped in simulations in which the context size, population size, or world dimensionality was increased. Although this is not surprising, in some cases the effects were very drastic, and highly dependent on the world's structure. Most noticeably, if the dimension values of objects in the world did not tend to be highly correlated with each other, then large context sizes or large world dimensionality resulted in enormous drops in communicative success. Moreover, even in successful simulations, the world structure sometimes influenced the rate of convergence. These patterns show that the simulation results do not easily scale up to larger systems, and thus must be treated cautiously. In particular, the world structure can have large consequences for whether a coherent lexicon will emerge and how long it will take. This points to the need for future models to choose their object representations or stimuli carefully and justify their choices.

Returning to the bigger picture, how exactly do these results relate to language evolution? To answer this, we need to revisit the hypothesis and clearly separate what exactly is being given a priori in this model, as opposed to what appears to be emerging (Steels, 2006). We started by asking whether self-organisation was able to explain how a population of hominins could have "invented" a lexicon from scratch. However, it's important to keep in mind that this hypothesis is framed within an implicitly substantial environmental and cognitive infrastructure. We have already seen that the environment that the agents are exposed to can play a crucial role in determining the outcomes of the simulations. The extent of the cognitive prerequisites have not, however, been substantially manipulated here. Agents are instead consistently endowed with unrealistically powerful and facilitating faculties (and in fact more so in some respects than some other models), including perfect word production and perception, powerful joint attention, limitless motivation for communication regardless of success, perfect and equal perception of objects, and perfect ability to use and interpret non-linguistic referential methods. All the simulation results of this model have done is to verify the internal consistency of the argument that, *given such abilities*, and under simple conditions, a lexicon could have emerged through a process of self-organisation, even if agents' categories had a prototype structure and they were always dealing with new objects. However, this work cannot address the question of whether the differences between the idealisations and the real phenomena are significant enough to give misleadingly optimistic results. In order to arrive at that, more work is needed, including integration with empirical experimental work with both humans and animals, as well as further modelling developments and explorations.

References

Belpaeme, T. (2002). *Factors influencing the origins of colour categories*. Unpublished doctoral dissertation, Vrje Universiteit Brussel.

Deacon, T. (1997). *The symbolic species: the coevolution of language and the brain*. New York: Norton.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*, 272–279.

Laskowski, C. (2006). *Prototype categorisation and the emergence of a lexicon in an infinite world*. Unpublished master's thesis, University of Edinburgh. (http://www.lel.ed.ac.uk/homes /cyp/dissertation/dissertation.pdf)

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. LLoyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Smith, A. D. M. (2003a). *Evolving communication through the inference of meaning*. Unpublished doctoral dissertation, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

Smith, A. D. M. (2003b). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, *9*, 559–574.

Steels, L. (1997). Constructing and sharing perceptual distinctions. In M. van Someren & G. Widmer (Eds.), *Proceedings of the European conference on machine learning* (pp. 4–13). Berlin: Springer-Verlag.

Steels, L. (2006). How to do experiments in artificial language evolution and why. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.), *The evolution of language: proceedings of the 6th international international conference* (pp. 323–332). London: World Scientific.