Could Navigation Be the Key to Language?

Dimitar Kazakov[⋆]
[⋆]Department of Computer Science,
University of York,
Heslington,
York YO10 5DD, UK,
kazakov@cs.york.ac.uk

Mark Bartlett[†]
[†]Department of Computer Science,
University of York,
Heslington,
York YO10 5DD, UK,
bartlett@cs.york.ac.uk

Abstract

This article analyses navigation and language parsing as two instances of the same abstract computation, and suggests that the tool needed may have evolved to serve the former task, and was then reused for the latter. Supporting evidence for the idea, based on the authors' concept of 'songline' navigation, is discussed in the context of current linguistic, psychological and neuroscience research. The discussion is concluded with an outline of a number of experiments that could shed further light on the subject.

1 Introduction

It is usually assumed that language as used by humans is inherently different from any other form of communication in other species, including primates. Facing this gap, and the intuitive expectation that several mechanisms would have to be already in place before language could play its role and provide any selective advantage, one can question what role (if any) evolution played in the emergence of language. Here we approach this issue by observing that a crucial feature of language, the ability to perform syntactic analysis and generate sentences from a set of grammar rules, can assist navigation, and suggesting that this ability may first have evolved for that purpose, and could be grounded in a general purpose neural circuit performing a certain class of abstract computations applicable across domains.

The article starts with a summary of current theories about the evolution and nature of the *language faculty*, the mechanism that allows us to acquire and use language. The following section discusses the use of computer simulations to model language evolution, and describes the Songlines model of navigation used in our simulations, and its implications on the potential link between the ways humans have evolved to handle navigation and language. This parallel, which we make from an abstract, computational perspective, is then compared with linguistic, psychological and neurological evidence supporting our suggestion. Further, the article describes the design for an experiment aimed at using navigation to detect the subject's ability to perform computation, analogous to context-free parsing. The last section summarises the main ideas and proposes a framework for comparative neuroimaging experiments that could help verify some of the claims made.

2 Evolution and Nature of Language Faculty

While the evolution of languages is an established idea nowadays, emphasised by the practice of grouping languages into genealogical trees of common descent (Pagel, 2000), the quest for the nature of the selective pressure that produced language has not ended in a consensus yet. Language has variously been suggested to emerge in order to provide information about the spacial aspects of the environment (O'Keefe and Nadel, 1978), maintain the social fabric of increasingly larger groups of hominids, e.g., to replace grooming and spread gossip (Dunbar, 1996; Power, 2000), etc. Byrne shows examples of primates using communication as a deception tool (Byrne, 1995), and it can easily be seen how full-fledged language would benefit this ability.

Another related question is the exact nature of the human language faculty and the extent to which it is innate. Chomsky claims that we are born with a Language Acquisition Device (LAD) (Chomsky, 1964), a complex blueprint, which sets its parameters when exposed to language. Marcus et al have reported that seven-month-old infants can learn to discriminate between the sentences of two different grammars (Marcus et al., 1999), reinforcing the belief that this ability is innate rather than acquired. While LAD has many supporters, the claim that it is a prerequisite to using language means that the coexistence of LAD and evolution as leading scientific paradigms is a somewhat uneasy one, as the notion that nature would have to put a potentially very complex tool in place before receiving a payoff contradicts the common wisdom that evolution usually advances in small steps, delivering immediate benefit. A 'macro-mutation' that would have produced the LAD all at once is extremely unlikely, and so would be a hypothesis that the components of LAD have been produced as a series of mutations, each amounting to 'genetic drift', that is, to a change that does not affect one's fitness.

Marcus et al's experiments were based on the familiarisation of subjects with sequences of syllables from an artificial grammar (e.g., both "ga na ga" and "li gi li" are instances of the general pattern ABA). During the test phase, novel spoken sequences, some of which violated the grammar, were played. A strong shift of attention towards the loudspeakers was judged as an acknowledgement of a perceived grammar violation. Importantly, the test sentences consisted entirely of new syllables. The study claims the infants were able to learn to recognise the general ABA pattern as different from ABB. The infants could also discriminate between the patterns AAB and ABB after being familiarised to either.

Marcus et al.'s interpretation of their results is that these are consistent with the infants' being able to "extract abstract algebra-like rules that represent relationships between placeholders (variables)", and that simple statistical learning relying on transitional probabilities cannot account for the experiments' outcome.[1]

Recently, many of the assumptions about the uniqueness of the human faculty of language have started to be questioned and experimentally tested (Hauser et al., 2002). A recent study (Fitch and Hauser, 2004) suggests there is a species (cotton-top tamarin monkeys) that can learn to recognise examples of *spoken* regular languages, but, unlike the human subjects in the studies, failed to learn a context-free language.[2] The results appear to extend the ground humans share with other species, and to point at the ability to handle CFGs as exclusively human.

In the reported work, Fitch and Hauser follow the familiarisation technique used by Marcus et al. (Marcus et al., 1999), but there are differences, for instance in the presence of overlap between syllables used in the training and testing phase. The RG ($(AB)^n$) and CFG (A^nB^n) used by Fitch and Hauser are more complex than those used on infants (ABB and ABA respectively), but still constrain n to 2 or 3 due to the memory limitations of tamarins.

Marcus et al. claim to have honed their methodology to eliminate the chance of having salient features in either grammar that would allow its recognition by statistical means. While following much of their precautions, and putting a considerable effort into eliminating any potentially salient non-grammatical features, Fitch and Hauser's work has been criticised for the use of different speakers for each of the A and B classes of symbols (syllables). If pitch is used as a feature, it is claimed, each of the languages studied would collapse to one example (for each grammar and value of n), thus reducing the experiment to one "about memory span and/or sensitivity to statistical deviations".

3 Songline Navigation

In recent years, there has been much research carried out in attempting to model the evolution of language through computer simulation. This research falls broadly into two classes, simulations in which language emerges in a single generation and simulations concerned with evolving a language over several generations.

In the former class, one of the most prominent researchers is Luc Steels. In his simulations (Steels, 1999), a population manages to arrive at a single, shared lexical language through participating in a series of 'language games'. In a language game, two agents discuss an object visible to both of them. If they can agree on a word (or set of words) to describe that object, then they both increase the weight they associate with that word/meaning pair. After many language games involving different pairs of agents, a shared global lexicon emerges.

Amongst those studying languages which are created over several generations, simulations presented by Kirby (Kirby, 2002) are amongst the most compelling, though Zuidema and Hogeweg (Zuidema and Hogeweg, 2000) and Oliphant and Batali (Oliphant and Batali, 1997) also present interesting results. In Kirby's simulations, a single agent attempts to express (resorting to invention if necessary) a subset of meanings sampled from a set of meanings, expressed in predicate calculus, while another agent attempts

[1] They however rely on the presence of such statistical learning, as observed by Saffran et al. (Saffran et al., 1996) to eliminate an alternative interpretation of their results.

[2] In the article, CFGs are discussed under the more general category of Phrase Structure Grammars.

to learn to speak based on the expressed meanings paired with the linguistic output of the first agent. The agent which listened is then required to speak in the same way as the first agent, while its output is learned by yet another agent. After thousands of cycles of this expression/induction behaviour, a grammar with the minimum number of necessary rules is seen to emerge and persist from generation to generation. Kirby attributes this to the 'linguistic bottleneck' that prevents the observation of all possible meaning/signal pairs by a single agent. Only compositional grammars can successfully pass through this bottleneck, as idiosyncratic phrases present in a grammar may fail to be expressed at some cycle and be lost from the language.

We have also chosen simulations as a means to study the evolution of language, but our approach, first outlined in (Kazakov and Bartlett, 2002), differs from all simulations mentioned above in several important aspects. Primarily, we see language as a tool to achieve some purpose. This means that we can consider issues such as when language will come to be used by a population, whereas other researchers have simply assumed that language is beneficial and sidestepped these issues. So far, our published work also differs by being unconcerned with the evolution of vocabulary: we assume that a shared lexicon has been fixed in the population by some means (for example through language games such as those used by Steels (Steels, 1999)), and concentrate on issues such as the mechanisms by which compositional language may have evolved and the types of environment in which it would be most beneficial (Bartlett and Kazakov, 2004).

Hamilton, and the following neo-Darwinist school of evolution, have developed a formal model of the phenomenon, and demonstrated on numerous examples that sharing among the individuals of a species is compatible with the concept of natural selection in the case of *kin selection* where help is directed to relatives in proportion to the degree of kinship (Hamilton, 1964; Dawkins, 1982). We assume that the mechanism of helping the poorer (hungrier, thirstier) agents has already been established in our simulated society (to simplify the matter, they are all considered equally related), and compare the benefits of sharing information about the location of the resource needed (food, water) between two agents with the case of help in kind, where part of the already collected resource changes hands.

A crucial observation, on which all work is based, is the fact that navigation can benefit from, and be based on a mental representation storing the route be-

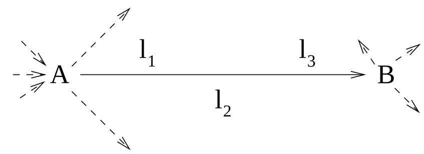

Figure 1: Navigation as a parsing task.

tween two locations A and B as a sequence of landmarks to be passed on the way. In the example in Fig. 1, to go to B, one has to be at (or go to) point A, then pass landmarks l_1, l_2 and l_3, in this order. This can be formally expressed as a rule:

$$\textit{goto}(B) \; :: \; \textit{goto}(A) \;\; l_1 \;\; l_2 \;\; l_3 \qquad (1)$$

To the experienced eye, this is, of course, a rule of a regular grammar, in which the start and end points of a route play the role of non-terminals (as there may be more than one way to reach and/or leave them), and landmarks are terminals. Therefore, tracing out (or following) a route between two points would amount to generating (resp. parsing) a sentence of a regular grammar (Kazakov and Bartlett, 2004). This representation was inspired by a socio-cultural phenomenon among the Australian Aborigines known as *Songlines*, a form of shared tribal memory, the knowledge of which is mandatory, and often secret, in which each song describes a landmark along a route, and the series of songs constitutes a sung map (Barwick and Marrett, 2003).

All that is needed to exchange a route between two agents using 'songlines' as their internal representation of the environment, is a shared lexicon of landmark names. Steels's experiments show that such a lexicon can easily be evolved from simple first principles (Steels, 1999). We assumed the existence of this lexicon, and sought to identify the types of environment in which sharing songlines outperforms selfish behaviour and sharing in kind. Among the factors studied were the abundance of the two types of resource modelled, and their volatility The results show language is particularly beneficial in the border zone delineated by the trade-off given by the simultaneous increase/decrease of resource availability and volatility. At one side of this line, there is too litle food and the one found disappears too quickly to be worth going back to; on the other side, food is sufficient to permit the survival of selfish agents. The area in the middle is where language appears to make the difference between survival and extinction (Kazakov and Bartlett, subm). In those areas, language outperforms both selfish behaviour and sharing previously accumulated resource.

4 Parallels between Navigation and Parsing

We have demonstrated that memorising and planning routes by an agent that describes a path as a sequence of landmarks (beacons) amounts to storing the rules of a regular language and generating/parsing its sentences. This is important: *if a regular language parser* (i.e., a Finite-State Automaton (FSA)) *could help navigation, it may first have evolved for this purpose*. Then only a relatively small change in the neural connections, possibly even caused by a single mutation, might have been required to make this parser available to the human brain speech circuitry. This compares favourably with the idea of macro-mutation, as described above. The idea of separately evolved needs for lexicon and grammar are also consistent with evidence that they are separated in the brain (Ullman, 2004).

We can consider now navigation and language parsing as two instances of the same abstract computation (involving strings of symbols) and enquire whether the way we perform these tasks would reflect that. Anyone interested in this question would be likely to look into existing models of the way in which syntax is grounded in the neural substrate. Ullman's recent model (Ullman, 2004) pinpoints several memory circuits in the brain, "a network of specific frontal, basal-ganglia, parietal and cerebellar structures", which support "the learning and execution of motor and cognitive skills, especially those involving sequences". The model separates, both neurophysiologically and conceptually, this so called procedural memory from the declarative memory storing information about facts and events, including the mental lexicon. The suggested common basis for the processing of verbal and non-verbal sequences is supported by other authors. Hoen et al. (Hoen et al., 2003) report that using non-verbal symbols (playing cards) to exersise the ability to reorder sequences in a predefined way ($123 \rightarrow 231$), helps patients improve their ability to understand a type of sentences that need the same transformation to have their constituents rearranged in the default order: "It was the cat^1 that the dog^2 chased^3" $\rightarrow$ "The dog^2 chased^3 the cat^1". Hauser, Chomsky and Fitch (Hauser et al., 2002) also draw a link between navigation and language, suggesting understanding efficient processing of language can help research in other domains, " such as spatial navigation and foraging, where problems of optimal search are relevant".

There are two ways in which the link between motor and verbal sequence processing may hold the key to the origins of syntax. One could conceive two coupled processes, (1) the need for 'songline' navigation providing selective pressure for the evolution of a parser, and, (2) the advantages of sharing 'songlines' promoting language. While it is very tempting to suggest that not only navigation may have selected for the evolution of a parser, but that it may have been the first topic of discussion to which this parser was applied, the second need not be the case. In fact, it is not easy to imagine how such a specialised language, possessing only nouns (or noun phrases) and the single verb *'to go'*, would have developed other parts of speech as a function of the navigational task. On the other hand, one can also consider another course of events, that the parser originally evolved to serve navigation as a means of internal representation and planning, and it was at a later stage that this parser (or its replica) became involved in communication. This resembles the idea of virtual mind machines processing whole classes of computationally homologous tasks, proposed by A. Sloman. This idea seems to be mirrored in neurological evidence about the parallel circuits of basal ganglia as performing analogous computations, applied to different sets of information from different domains (Ullman, 2004).

5 From Regular to Context-Free Languages

So far, we have only demonstrated that navigation could have created the need to represent and process regular languages. However, RGs are not sufficient to describe the seemingly simple task of going somewhere and returning back using the same way. In terms of 'songlines', the landmarks passed on the round trip would spell out a palindrome (e.g., $abcba$ or $aabbaa$). It is known that a parser that can recognise a palindrome, can handle any CFL. A CFL parser is usually modelled as a push-down automaton, consisting of a FSA (i.e., a RL parser) and memory (e.g. stack). Regardless of whether this task created the original need for CF parsing or not, it can be used in an experiment involving navigation, rather than responses to speech, which could help avoid the criticism to Fitch and Hauser's work (Fitch and Hauser, 2004), while still assessing the subject's ability to learn syntax.

One could consider two classes of tasks. In the first, the subject would have to learn to navigate be-

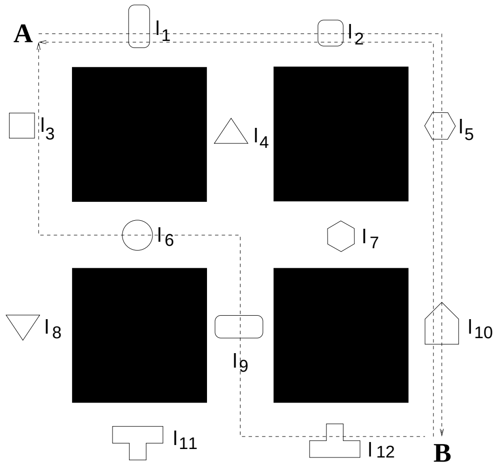

Figure 2: Regular vs Context-Free Songlines.

tween two locations, e.g., A and B in figure 2, collecting a reward each time either location is reached. Assuming landmarks are used to memorise the paths between A and B as suggested above (Kazakov and Bartlett, 2004), successful navigation would amount to learning a regular grammar, e.g.:

$$\textit{Reward} \rightarrow B \quad (2)$$
$$B \rightarrow A\ l_1\ l_2\ l_5\ l_{10} \quad (3)$$
$$\textit{Reward} \rightarrow A \quad (4)$$
$$A \rightarrow B\ l_{12}\ l_9\ l_6\ l_3 \quad (5)$$

The second experiment modifies the above setting by extending the reward given in location A to a more significant one, provided the subject went from A to B and back using the same path. Navigation based on a regular grammar with alternative routes will fail to collect the extended reward most of the time, e.g., there are 18^2 ways of going from A to B and back in figure 2, but only 18 of these will bring the maximum reward. However, a simple context-free grammar will suffice:

$$\textit{Reward}_1 \rightarrow A\ \textit{Reward}_2\ A \quad (6)$$
$$\textit{Reward}_2 \rightarrow X \quad (7)$$
$$X \rightarrow \textit{Landmark}\ X\ \textit{Landmark} \quad (8)$$
$$X \rightarrow B \quad (9)$$
$$\textit{Landmark} \rightarrow l_i,\ 1 \leq i \leq 12 \quad (10)$$

Using another string of salient features, such as turns and distances will not change the need to reverse that string to navigate back (but will assume the ability to transform a left turn to a right one and vice versa). One can use obstacles to guarantee that no complete path used in training is available in the test phase and vice versa.

Classical Reinforcement Learning (RL) (Sutton and Barto, 1998) cannot account for learning this navigational behaviour. With consecutive phases of (random walk) exploration and exploitation, RL will assign in average equal rewards to all A *to* B, resp. B *to* A paths; if the agent alternates between exploration and exploitation, and gradually increases the latter, the earlier an A *to* B, resp. B *to* A path is discovered, the more likely it is to be subsequently reinforced, and given preference in the long term.

The above setting would avoid issues stemming from the much greater importance speech has for humans and put them on a more equal ground with other species. A confirmation of Fitch and Hauser's conclusions that possessing a CF parser is a distinctly human feature would raise the question whether this did not initially evolve to serve non-linguistic purposes, such as navigation, and study the circumstances that made this new feature evolutionary beneficial. For instance, taking the same way back home may reduce the risk of encountering unexpected dangers or help estimate the time needed to return back. On the other hand, a shared need for navigation among species would be consistent with them sharing the ability to learn regular languages.

6 Discussion

The main ideas in this article can be summarised as follows:

1. The ability to hangle regular grammar, a critical step on the road to human language, may originally have evolved to assist navigation.

2. The shared need for navigation should be mirrored in the ability of other species to learn regular languages.

3. Navigation and language parsing are two instances of the same abstract computation, and the way they are grounded may reflect that.

4. The need for context-free grammars, typical for human languages, could have originated in navigation.

The range of indirect evidence for the above statements suggests the idea of using neuroimaging to compare the brain activity between tasks correspongging to regular and context-free languages for navigation on one hand, and language, on the other. An exciting, but yet unconfirmed possibility is that the patterns of activation for navigation and language would be similar for the same class of languages, but processing a different class of language would result in distinguishable differences even for the same type of task.

References

M. Bartlett and D. Kazakov. The role of environment structure in multi-agent simulations of language evolution. In *Proceedings of the Fourth Symposium on Adaptive Agents and Multi-Agent Systems (AAMAS-4)*, Leeds, 2004. AISB Convention.

L. Barwick and A Marrett. *Currency Companion to Dance and Music in Australia*, chapter Aboriginal Traditions, pages 26–28. Currency Press, 2003.

Richard W. Byrne. *The Thinking Ape: Evolutionary Origins of Intelligence*. Oxford University Press, 1995.

Noam Chomsky. *Current Issues in Linguistic Theory*. Mouton, The Hague, 1964.

R. Dawkins. *The Extended Phenotype*. Oxford University Press, 1982.

Robin Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1996.

M. T. Fitch and M. D. Hauser. Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303, 2004.

W. D. Hamilton. The genetic evolution of social behaviour I, II. *Journal of Theoretical Biology*, 7: 1–52, 1964.

M. D. Hauser, N. Chomsky, and M. T. Fitch. The faculty of language. *Sci.*, 298, 2002.

M. Hoen, M. Golembiowski, E. Guyot, V. Deprez, D. Caplan, and P. F. Dominey. Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport*, 14, 2003.

D. Kazakov and M. Bartlett. A multi-agent simulation of the evolution of language. In *Proceedings of Information Society Conference IS'2002*, pages 39–41, Ljubljana, Slovenia, Oct 2002. Morgan Kaufmann.

Dimitar Kazakov and Mark Bartlett. Cooperative navigation and the faculty of language. *Applied Artificial Intelligence*, 19, 2004.

Dimitar Kazakov and Mark Bartlett. Benefits of using language to share information in a dynamic environment. Submitted to Connection Science (special issue on Evolution of Language), subm.

S. Kirby. Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, 2002.

G. F. Marcus, S. Vijayan, S. Bandi Rao, and P. M. Vishton. Rule learning by seven-month-old infants. *Science*, 283, 1999.

J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Oxford Univ. Press, 1978.

M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *The newsletter of the Center for Research in Language*, 11(1), 1997.

M. Pagel. The history, rate and pattern of world linguistic evolution. In *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, Cambridge, 2000. Cambridge University Press.

C. Power. *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, chapter Secret language use at female initiation: Bounding gossiping communities. Cambridge University Press, Cambridge, 2000.

J. Saffran, R. Aslin, and E. Newport. Statistical learning by 8-month old infants. *Science*, 274, 1996.

Luc Steels. The spontaneous self-organization of an adaptive language. In Koichi Furukawa, Donald Michie, and Stephen Muggleton, editors, *Machine Intelligence 15*, pages 205–224. Oxford University Press, 1999.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

M.T. Ullman. Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 2004.

W. H. Zuidema and P. Hogeweg. Selective advantages of syntactic language - a model study. In *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*, pages 577–582, Hillsdale, USA, 2000. Lawrence Erlbaum Associates.