Modelling the Emergence of Case

Joanna Moy and Suresh Manandhar
Department of Computer Science
University of York

http://www.cs.york.ac.uk/~joanna

1 Introduction

A further investigation into the role of linguistic evolution as an alternative to biological evolution in the emergence of syntax is presented. This follows on from the idea that languages themselves are evolving entities, which adapt to be easily acquired by the human learner. It has already been shown that it is possible for the rudimentary elements of syntax, i.e. compositionality and recursiveness, to emerge in a population of language learning agents without any specific linguistic knowledge and in the absence of selection for communicative ability[8],[7]. Attempts are made here to extend this framework to cover another important aspect of natural language: the use of a case system to make distinctions between thematic roles.

1.1 The Origins of Human Language

Human language seems to be both qualitatively and quantitatively unlike any other form of animal communication. Unsurprisingly, then, the origins of this uniquely human ability are the subject of much debate, particularly the question of nature versus nurture. Is language innate? Or must it be "discovered" anew by each new human learner? Or is the real answer somewhere between these two extremes?

That we are in some respects predisposed to language is beyond doubt. Terrence Deacon suggests that human beings manifest an extensive array of perceptual, motor, learning, and even emotional predispositions towards the learning of language[5]. For example, the positioning of the human larynx is significantly lower in the throat than that of other primates, and we have a greater degree voluntary control over our respiratory function, enabling the long slow exhalations necessary for the production of speech. Although cognitive adaptations to language are less easy to identify than physiological ones, it is worth noting that "all normal children, raised in normal social environments, inevitably learn their local language, whereas other species, even when raised and taught in this same environment, do not". Attempts to teach symbolic communication to non-human primates have produced some interesting results, but it is clear that they do not have the same facility and natural aptitude for language that we do. Language emerges in all normal children all over the world at approximately the same age, no matter what culture they are growing up in and what language they are acquiring[1]. Furthermore, Briscoe notes that "failure [to acquire language] appears to correlate more with genetic defects or with an almost complete lack of linguistic input during the critical period, than with measures of general intelligence or the quality or informativeness of the learning environment"[3].

It is clear that there is "something special about human brains that enables us to do with ease what no other species can do even minimally without intense effort and remarkably insightful training"[5].

So perhaps the real question is this: what aspects of the ability to learn language are innate, and what form do they take?

1.2 Principles and Parameters

The Chomskyan position is generally stated to be that children are born with a large amount of innate knowledge regarding the structure of language (see Culicover[4] for review). This premise is based on the observation that we do not learn to speak by merely repeating utterances that we have heard, but instead internalize a set of rules for the construction of novel utterances, and yet the language to which children are exposed is not sufficient to make these rules explicit. If the structure of language is underspecified, then children must need additional guidance to successfully acquire the rules. In his earlier work, Chomsky suggests that this might take the form of information about linguistic universals, i.e. knowledge of the general outline of language, a kind of "blueprint". Acquisition simply becomes a problem of choosing the appropriate language from within the space of possible languages, for which the existence of a specific cognitive module or Language Acquisition Device is postulated. Chomsky's more recent work on the Principles and Parameters framework takes this idea further still, in proposing a Universal Grammar which constrains the space of possible languages even more tightly. Language acquisition is reduced to the setting of a "finite number of finite-valued parameters".

There are many objections to the Chomskyan theory. Among them is the idea that a Language Acquisition Device is impossible because it could never have arisen by Darwinian natural selection. Instead it must have required some kind of macro-mutation in order to come into existence. The debate rages on this issue, as Chomsky himself believes that natural selection cannot account for his theory, and wishes instead to invoke "alternative physical principles". However, Stephen Pinker[11], also strongly in favour of the nativist view of language acquisition, (although of a slightly different flavour), argues that natural selection, by virtue of its directedness, is the only process that *can* account for such a highly specialized organ. Another area of controversy is whether or not child-language learning data actually fits the patterns predicted by the theory. Aitchison[1] suggests that if the Principles and Parameters model is correct, then children should be aware of the innate constraints on the form of language and should never make an utterance that is outside the scope of all natural languages. Also, their acquisition should proceed in dramatic steps as the appropriate parameter settings are acquired. Yet there seems to be plentiful child-language data which suggests that this is not the case[13].

1.3 Language as an Adaptive System

An alternative to the Universal Grammar is the idea that, rather than having brains that are pre-wired with quite specific information about language, we are instead equipped with puzzle-solving apparatus that enables us to successfully process linguistic data as we come across it[1]. The stronger form of this hypothesis is that children possess language-specific puzzle-solvers, whereas in the weaker form the mechanisms used to acquire language are simply a facet of the general intelligence of human beings.

A yet more radical stance suggests that Chomsky has inverted cause and effect. Rather than suggesting that human brains have evolved to be better able to acquire language, could it be that languages themselves evolve to be more easily acquired by human brains? Can languages themselves be viewed as evolving entities, which have adapted to suit the cognitive processes present in the human mind? When, as Chomsky observed, children quickly master the basics of grammar despite the paucity of the input stimuli from which they learn, one could perceive them to be making "lucky guesses" about grammar and syntax and the way words work together. These might be all too easy to attribute to some innate language-specific knowledge. However, Terrence Deacon[5] believes that learning is occurring by trial and error, but that a very large proportion of the guesses children make are correct – not because they possess innate knowledge about language, but because the languages they are learning have evolved in such a way as to fit in with the guesses that are likely to be made.

It is possible to view the relationship between a language and its speaker as a similar to that between a virus and its host. Just as the viral DNA uses its host cell in order to reproduce, the information encapsulated in a grammar of a language becomes integrated into the machinery of the human brain, and uses this to reproduce.[1] The metaphor can be taken further still by the observation that although a common language may link a social group, the *internal* grammar of each of those speakers is unlikely to be identical, but subject to variation – thus the language of their community is more like a collection of similar but not identical languages. And as a result of this variation, languages are able to evolve with respect to the selection pressures around them, just as variation within a viral gene-pool drives evolution.

The selection pressures in question are the biases of human learners. In making their "lucky guesses" about the way in which words work together, children actually neglect a large proportion of the hypothesis space – they fail to explore the full range of possible ways of organising words. Thus a language that organises its words in a way that falls outside the "lucky guesses" of children will not be easily learnt (and will be less likely to pass on to the next generation), whereas a language for the whom the "lucky guesses" are correct guesses will be much more easily acquired.

The emergence of simple languages and compositional structures as a result of this process has been demonstrated using computer simulations[2],[8]. In particular, Kirby has done a series of experiments in which he has demonstrated the evolution of syntax in a population of agents equipped with a set of simple grammar learning heuristics, but which are not themselves subjected to any selective pressures, [6],[7],[8],[9], showing that it is possible for a language to evolve independently of its "host".

Kirby's work incorporates what he refers to as the "iterated learning model" where agents learn observationally from the behaviour of others. Agents in the simulation are all identical, and there is no selection for communicative ability. Learners attempt to build a grammar by extracting regularities from the utterances they hear, with the result that they quickly converge on a very tidy, minimal, fully compositional grammar to express the meaning space required. Kirby explains this as a consequence of the "dynamics of language transmission", that is the fact that an agent cannot hope to sample utterances for the whole of the meaning space during its lifetime, and thus a language in which the meaning of a string can easily be predicted from its structure will stand a much better chance of being propagated from one generation to the next. Initially there is no such language, but if similarities between strings with similar meanings should occur by chance, and if the appropriate generalizations are made, structure within the language will be selected for.

The work described in this paper is a further investigation of the role of linguistic evolution, (as opposed to biological evolution), in the emergence of syntax. A key feature of the grammars that emerge from Kirby's simulations is that they use word order to specify meaning distinctions. However, it is worth noting that in many natural languages, meaning distinctions are not wholly specified by word order – even in English, some freedom of word order is allowed, and other languages allow much more. This is generally accompanied by a much richer case system than that found in English. The purpose of this work is to see if it is possible to use freedom of word order as a driving force towards the emergence of a case system.

---

[1] A key difference however, is that this relationship is symbiotic (unlike that between virus and host) because both host and language benefit from the other.

2 Modelling a Primitive Case System

The current work was carried out using an iterated learning model of language evolution based on that described by Kirby[7], to which the reader is referred for the details of the implementation.[2] At any given moment in time there exist two agents in the system, a learner and a speaker. The speaker is required to produce an utterance to express a meaning drawn from a simple meaning space. The utterance is made up of a string, composed of a number of alphabetic characters (intended to represent the irreducible phonemes of a language), plus a representation of that meaning. The meaning space is composed of "who did what to whom" type ideas, drawn from a set of five possible participants, plus five actions each requiring an agent and a patient. The speaker selects an appropriate string by consulting its own internal grammar, which, as in Kirby[7] is represented as context-free grammar enriched with simple semantics: non-terminal symbols have a single argument associated with them which conveys semantic information. If the agent's grammar does not specify an appropriate string for the required meaning, the agent will invent a new string (or a new part of a string) with a predefined probability (here simply set to 1). As the initial population has no grammar, this will be the only way that utterances can be produced in the early stages of the simulation. When presented with the speaker's utterance, the learning agent first attempts to see if it is covered by its own grammar, by attempting to parse the string. If the parse is successful and the correct meaning is returned, then the learner does nothing further. If not, the agent attempts to learn from the input according to the simple grammar learning heuristics described below. This process is repeated 100 times, at which point the current speaker is removed from the simulation, the current learner becomes the new speaker, and a new learner with an empty grammar is "born".

The grammar learning heuristics consist of two basic phases. The first is a simple incorporation step, whereby the utterance is added to the agent's grammar as a simple grammar rule of the form `s/[meaning]-->string`. The second stage is to make generalizations between this utterance and the others it has heard. This involves selecting pairs of rules and seeking to create a rule that will subsume them both. To this end, there are two operations available to the agent:

- If rules A and B differ only by non-terminals X and Y, and if changing Y to X would make them identical, then rule B is removed, and all other instances of Y in the grammar are changed to X.
- If the semantics of rules A and B differ by the value of a single element whose meanings are a and b, and their strings differ by substrings α and β, a and b are replaced by a variable x, and α and β are replaced by a non-terminal, N whose meaning is x. New production rules are created from N to strings α and β with meanings a and b respectively (effectively attributing the differences between the two meanings to the differences in the strings).

As in Kirby's system, the language spoken by the population of the simulation evolves over a number of generations from a simple vocabulary driven language, where each meaning is represented by an idiosyncratic string with no internal structure, to a fully compositional language, in which the meaning of the string is derived from the meaning of its parts, and the way they are assembled. In particular, separate syntactic categories for nouns and verbs emerge, and there is use of word order to distinguish between thematic roles.

Free word order languages do not emerge, and nor does the use of inflection to specify the distinction between thematic roles. This is not entirely surprising, given the nature of the heuristics that are used in grammar induction. These heuristics work by comparing strings and looking for common

---

[2] Although this system does not deal with recursion as Kirby's does.

prefixes and suffixes. When the parts of the strings which are the same have been identified, those which are different can be attributed to differences in the meanings of the two strings (as described above). Thus if presented with the strings *abcdef* meaning “John loves Mary”, and *abcdgh* meaning “John loves Kate”, the grammar inducer would identify the common prefix *abcd*, whilst noting that the final sections of the two strings differ. Thus the difference in meaning would be ascribed to this, resulting in the conclusion that *ef* means “Mary”, whilst *gh* means “Kate”. Suppose however, that the second string had been *ghabcd*, as might occur in a language that allows freedom of word order. This shares neither a common prefix nor suffix with the string *abcdef*, so the current grammar inducer would fail to notice any similarity between the two. As a result it would not pick out the relevant differences either. Thus if chance regularities between strings did occur such that a free word order language might be induced, the current grammar inducer would not be able to induce it.

However, natural languages do not tend to exhibit such rigid word order as those emerging from the simulation. Even English, which has a relatively strict ordering, allows a small degree of word order freedom, for example when the speaker wishes to topicalize the object. Other languages, such as German allow a lot more, and still others exist which allow almost complete freedom of word order. Clearly, in such languages, it is no longer possible to use word order to distinguish between thematic roles: if the language allows both SVO and OVS sentences, for example, and the string *johnlovesmary* is heard, how is the hearer to distinguish between the two possible meanings “John loves Mary” and “Mary loves John”? Instead, inflection is commonly used – different forms of the nouns *john* and *mary* to determine their case, i.e. whether they are subject or object of the sentence. Thus the two possible meanings can be distinguished by the form of each noun used.

It is noted that in the original system, occasionally a grammar with two noun categories rather than one emerges, so that agents are effectively using different subject and object versions of each noun, despite having a fixed word order. The current work is an attempt to see if it is possible to create a selective pressure for languages of this type. Will a degree of word order freedom cause the emergence of case for distinguishing thematic roles once word order is no longer a reliable cue? A number of changes were made:

- Firstly a function was added to “shuffle” the utterance made by the speaking agent with a fixed probability p, i.e. the parts of the utterance are re-ordered randomly. In the early stages of the simulation, where the grammar is composed entirely of idiosyncratic strings, this will have little effect, whereas later, once separate syntactic categories have begun to evolve, it will have the result of re-ordering the subject, object and verb components of the sentence. This is intended to model the occasional “mistake” or use of word order to provide emphasis by the speaker. Once an utterance with an alternative word order has been made, if it is incorporated into the learner’s grammar, the rules producing that word order may well be used again, when that agent becomes a speaker. Thus it may be propagated down the generations. It is hoped, that by generating potentially ambiguous word orders, this will act as a driving force towards distinguishable subject and object versions of each noun, i.e. a primitive case system.

- Secondly it was necessary to enable the system to make use of multiple rules with different word orders. As the original system is deterministic, given a choice between two possible rules to generate an utterance for a given meaning, the same one will be used every time. Clearly this will not allow the propagation of alternative word orders, so the parser/generator was changed to work on a probabilistic basis. A count was associated with each rule in the grammar, according to the number of times that rule has been used. When generating utterances, if there is more than one rule that can produce a string with the required meaning, the probability of choosing a given rule is weighted according to its count.

- Finally, the fact that ambiguous utterances will be perfectly tolerated was addressed. Tolerance of ambiguity works against the emergence of case because there is no need for a means of disambiguation to become established. In order to overcome this, a *distinguishability* flag was added, which could be set to either 0 or 1 depending on whether or not the agents require utterances to be distinguishable. When the flag is set to 0, the system behaves as it did in its original incarnation, and is completely tolerant of ambiguity. The learner simply looks to see if it can parse the string to give the correct meaning, and if so, it is satisfied, and does nothing further. If the string cannot be parsed at all, or if the meaning returned is incorrect, it adds it to its grammar associated with the intended meaning. This means that a single string can be associated with any number of meanings. However, when the flag is set to 1, agents assume that each utterance is completely unambiguous. When presented with an utterance-meaning pair by the speaker, the learner again tries to parse that utterance, but this time is satisfied if it can parse the string to give *any meaning*. If a parse is possible, it assumes it already knows that utterance, *even if the meaning returned is incorrect*. It only incorporates a string into its grammar if it cannot parse it at all. So, according to the revised rules, if the speaker agent only has a single noun category, used to express both subjects and objects, and it then shuffles its output in such a way that the subject and object are inverted, the learner will never acquire the rule required to generate this sentence. However, if the speaker has two separate noun categories, one to represent each of subjects and objects, the learner will be able to acquire this rule. This should result in a selection pressure for separate subject and object noun categories.

3 Results

The results obtained from these changes to the simulation appear very promising at first. Running the system with p set to 0.01 and the distinguishability flag set to 1, the resulting grammars can be subdivided into two types. The first (Type A) has two separate noun categories for expressing subject and object and these grammars generally allow a large range of the possible word orders.[3] The second type of grammar (Type B) has only one noun category, used for both subject and object, but a more restricted range of word orders. However, word order is not entirely fixed: in general, approximately half of the possible word orders are represented. The set of possible word orders can be subdivided into pairs, which are identical but for the fact that the subject and the object have been transposed. If subject and object forms of any given noun are identical, this makes it impossible to determine which of the two rules is being applied and thus distinguish which noun is the subject and which is the object. Therefore, orderings from these pairs are mutually exclusive in the Type B grammars. If a sentence is made up of two nouns plus a verb and no other characters, this allows a total of six permutations. Generally three of these are expressed, without any loss of distinguishability of subject and object. For example, in the case where the word order SOV is allowed, SVO will not be, because this would cause confusion. Other rules such VSO can be perfectly easily distinguished from this however, due to the different positioning of the verb.

A typical Type B grammar looks like this:[4]

---

[3] Interestingly, although two separate noun categories have emerged, and within any given rule one of them is used for the subject of the sentence and one for the object, there is no consistency *between* rules: in the grammar above, three of the top level rules for building a sentence use category 1 to represent the subject and category 3 to represent the object, whilst one rule uses category 3 for subject and category 1 for object.

[4] Where P, X and Y are variables over predicates, subjects and objects, respectively.

```
s/[P, X, Y] --> [3/Y, 2/P, 3/X]
s/[P, X, Y] --> [2/P, 3/X, 3/Y]
s/[P, X, Y] --> [3/X, 3/Y, 2/P]
3/john   --> [q]
3/mary   --> [t]
3/pete   --> [u, f]
3/anna   --> [z, e]
3/kath   --> [r]
2/loves  --> [c]
2/hates  --> [r, a]
2/adores --> [i, t]
2/kisses --> [i]
2/sees   --> [m, j, g]
```

It can be seen that this grammar exhibits three of the six possible word orders for sentences sentences made up of two nouns, plus one verb: OVS, VSO and SOV. From each of the mutually exclusive pairs, SOV-OSV, VSO-OSV and SVO-OVS, only one is present. And of the three that do occur, the position of the verb makes it quite clear which rule is being applied, and thus it is possible to differentiate which noun is subject and which is object. This can be contrasted with the typical Type A grammar shown below. This one exhibits four possible word orders, OVS, SOV, VSO and SVO:

```
s/[P, X, Y] --> [3/Y, 4/P, 1/X]
s/[P, X, Y] --> [1/X, 3/Y, 4/P]
s/[P, X, Y] --> [4/P, 3/X, 1/Y]
s/[P, X, Y] --> [1/X, 4/P, 3/Y]
1/john   --> [i]
1/mary   --> [f, z, x]
1/pete   --> [h, n, v]
1/anna   --> [j]
1/kath   --> [y]
3/john   --> [a, t]
3/mary   --> [d]
3/pete   --> [l]
3/anna   --> [i, u]
3/kath   --> [q]
4/loves  --> [c]
4/hates  --> [k, h, k]
4/adores --> [h, i, x]
4/kisses --> [f]
4/sees   --> [l]
```

It should be noted that although only four of the possible six word orders are displayed in this grammar, it includes both members of one of the pairs of orderings which are mutually exclusive in the Type B grammar: OVS and SVO. This is made possible by the existence of two noun categories. Thus the meaning “John loves Mary”, using the OVS rule would be *dci* which is perfectly distinguishable from “Mary loves John” using the SVO rule, which is *fzxcat*, even though both sentences involve a word for “Mary” followed by a word for “loves” followed by a word for “John”.

Type A grammars seem to appear at a slightly higher frequency that Type B: 56% of the runs carried out exhibited a Type A grammar, and 44% a Type B. This is in contrast to the occurrence of separate noun categories in the original system, which only happened in approximately 24% of the runs. It would appear that the attempts to create a selective pressure for some form of case marking have been successful.

However, further investigation shows that things are not quite as they may seem. Re-running the simulation with p set to 0 (no shuffling) and the distinguishability flag also set to 0 (ambiguity is tolerated), i.e. as in the original system, only using the probabilistic parser resulted in 2-noun category grammars in 60% of the runs. There is also occasional spontaneous re-ordering of the sentence structure.

Returning to a p value of 0.01 (1% of utterances are shuffled) but with the distinguishability flag still set to 0 (ambiguity still tolerated), the number of grammars exhibiting case is reduced slightly to about 54% which is close to the value found with the distinguishability flag set to 1. However, the results are still quite different. For both grammars with case and grammars without case, all of the possible word order permutations are at some point added to the grammar. Whether or not they survive and are propagated to future generations is purely a matter of chance: with the probability of shuffling set at 0.01, it is unlikely that a single speaker will spontaneously produced the same “shuffled” sentence more than once, therefore the count associated with the rule specifying that particular word order will be very low. If that rule is chosen, it will be incorporated in the grammar of the next learner, but again it is unlikely to be chosen more than once, so again will result in a rule with a low probability of being chosen. Therefore most of the alternative word orders are quickly lost. However over the course of several thousand generations, rules for alternative word orders are learnt and propagated, and gradually many of the possible permutations are added to the grammar.[5] The pattern of word order restriction seen in the Type B grammar does not occur.

4 Conclusions and Further Work

The current work has shown that it is possible for a primitive case system (i.e. separate noun forms to represent subject and object) to emerge from a population of learners equipped with a simple learning algorithm for grammar induction, but no language specific knowledge. However, it would appear that it is not variability of word order that drives this emergence, as initially predicted. Nonetheless, once case has emerged, it certainly facilitates the use of alternative word orders, by helping to resolve some of the ambiguities that they may generate. This makes it possible for a language with a case system to support a wider range of word orders than one without. However, in those populations in which case does not emerge, the language need not be restricted to just a single word order, but to a very specific subset of those available: those which are easily distinguishable from each other by the positioning of other elements of the utterance.

This is an interesting result because such a pattern of restricted word order is quite unlike natural language, where it is generally the case that word order is either fairly strict, as in English, or allows a full range of word orders, e.g. languages such as Turkish and Serbo-Croatian[12]. Even when languages do exhibit a restricted set of word orders, such as Italian, which is predominantly SVO, but in which OVS, VSO and VOS are also relatively commonplace, this subset often includes pairs

---

[5] Generally no more than 5 different word orders are expressed at any one time however. This is because each alternative word order is added at the expense of the others – for any new order to get high enough counts to have a reasonable chance of survival, the others must have their counts reduced. If the count on any one of these is reduced too much, then it becomes likely that it will die out instead. Importantly though, there is no restriction on *which* word orders can appear together.

of word orders for which other cues are required to specify subject and object. Interestingly, Italian does not have a particularly rich case structure, and these alternative word orders are generally only allowed when the meaning can be disambiguated from the context.

As an extension to the present work, the "distinguishability flag" used to prevent agents from tolerating ambiguity has been removed, as this is clearly unrealistic. Instead, an additional phase has been added to the simulation in which teacher and learner converse with each other. No additional grammar rules are learnt during this phase, but if a string uttered by the conversational partner is parsed to give an incorrect meaning, then the rules used in that parse will be penalised, resulting in a decreased likelihood of those rules being used in the future. Early results from these simulations seem to show that in the case of grammars with only a single noun category, there is a much higher tendency towards a single dominant word order. Presumably this is because both versions of an ambiguous word order have an equal likelihood of being penalised, and are therefore equally likely to have their counts reduced to a very low level, and ultimately to disappear, as opposed to the situation where one rule gets established at the other's expense. However, this still does not result in a truly natural language-like distribution of word orders.

The results of this study suggest a number of possible future directions. First and foremost, an investigation of the mechanism by which the probabilistic parser drives agents to evolve case is required. Is the related to the occasional spontaneous re-ordering of sentences? And how does this spontaneous re-ordering occur? It was expected that the emergence of case would be driven by the need to disambiguate between possibly conflicting word orders, and yet adding this pressure does not appear to drive the system any further towards case-based languages, if anything it seems to slightly reduce their likelihood.

Secondly, it would appear that in natural language also, mechanisms other than the need to disambiguate are at work in the development of case. Otherwise there would be no reason why languages should exhibit completely fixed word order, and patterns such as that found in Italian would be somewhat unlikely. This may perhaps be related to sentence processing demands: Lupyan and Christiansen[10] have done studies using simple recurrent networks, which demonstrate that certain word orders are easier to learn than others. In particular, SVO and OSV are more readily acquired than SOV. In this case, the addition of case markings aids acquisition. This is in keeping with the fact that in natural language, SOV languages generally exhibit case markings, whereas most caseless languages are SVO or VSO. Therefore it appears that case markings may originally appear in order to facilitate the acquisition of fixed word order languages whose underlying word order is difficult to learn, but that their existence might enable more freedom in the word order used. In the current system, all possible word orders are equally easy to learn and equally likely. It would be interesting, therefore, to try and reproduce the effects of word order on learnability, and to see if this results in patterns of word order restriction more akin to those found in natural language.

Finally, the "case system" that emerges in the current study is far from representative of the type of case found in natural languages, which normally takes the form of inflectional affixes. In the present results, subject and object forms of the same noun are completely unrelated. Whilst this might occur for certain irregular wordforms that are very frequently used in a given language (such as "we" and "us"), it is not the norm. A planned follow-up, therefore, is to extend the current system to generalize across any chance regularities that may occur between subject and object forms of a given noun, or across different nouns of the same case, in the hope that a truly inflectional case system may be derived.

In conclusion then, although follow-up work is clearly required, the present study has demonstrated the emergence of a primitive form of case within the evolutionary framework described above.

References

[1] Jean Aitchison. *The Articulate Mammal.* Routledge, fourth edition, 1998.

[2] John Batali. Computational simulations of the emergence of grammar. In James Hurford, Chris Knight, and Michael Studdert-Kennedy, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge University Press, 1998.

[3] Ted Briscoe. Grammatical acquisition and linguistic selection. In E. J. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, 1999.

[4] Peter W. Culicover. *Principles and Parameters: An Introduction to Syntactic Theory*. Oxford University Press, 1997.

[5] Terrence Deacon. *The Symbolic Species: The Co-evolution of Language and the Human Brain*. Penguin Books, 1997.

[6] Simon Kirby. Language evolution without natural selection: From vocabulary to syntax in a population of learners. Technical report, University of Edinburgh, 1998. EOPL-98-1.

[7] Simon Kirby. Learning, bottlenecks and the evolution of recursive syntax. In Ted Briscoe, editor, *Linguisitic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, 1999.

[8] Simon Kirby. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Chris Knight, Michael Studdert-Kennedy, and James Hurford, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, 2000.

[9] Simon Kirby. Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions of Evolutionary Computation*, 5(2):102–110, 2001.

[10] Gary Lupyan and Morten H. Christiansen. Case, word order, and language learnability: Insights from connectionist modeling. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, to appear.

[11] Steven Pinker. *The Language Instinct.* Penguin, 1994.

[12] Dan I. Slobin and Thomas G. Bever. Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12:229–265, 1982.

[13] B Wilson and A M Peters. What are you cookin' on a hot? *Language*, 64:249–73, 1988.