The Language Mosaic and its Evolution, James R Hurford,

The human capacity for language and the structures of individual languages can best be understood from an evolutionary perspective. Biological steps toward language-readiness involved preadaptations for modern phonetics, syntax, semantics and pragmatics. Once humans were language-ready, ever more complex language systems could grow, relatively fast, by cultural transmission, generation after generation.

0.1 What language evolution research is about

It is natural to ask fact-demanding questions about the evolution of language, such as `Did *Homo erectus* use syntactic language?´, `When did relative clauses appear?', and `What language was spoken by the first *Homo sapiens sapiens* who migrated out of Africa?'. I believe, however, that study in the evolution of language will not yield answers to such questions in the near future. Therefore, finding answers to such empirical-in-principle questions cannot be the purpose of language evolution research. The goal is, rather, to explain the present.

Evolutionary linguistics does not appeal to an apparatus of postulated abstract principles specific to the subject. Language is embedded in human psychology and society, and is ultimately governed by the same physical principles as galaxies and mesons. Not being physicists, or even chemists, we can take for granted what those scientists give us. The basic linguistic facts needing explanation are these: there are thousands of different languages spoken in the world; these languages have extremely complex structure; and humans uniquely (barring a tiny minority of pathological cases) can learn any of these languages. Modern mainstream linguistics has ignored the single most promising dimension of explanation, the evolutionary dimension.

Any evolutionary story of how complex organisms arose must now be consistent with what we know about the behaviour of molecules. The evolutionary story must also be consistent with knowledge from another new and independent body of theory. This work emphasizes that the environment in which natural selection operates is characterized by mathematical principles, which constrain the range of attractor states into which evolution can gravitate.

The advent of social behaviour necessitates even more complex methods of analysis, many not susceptible to mathematical modelling, due to the highly nonlinear nature of advanced biosocial systems. With plasticity (especially learning) and advanced social behaviour comes the possibility of culture, and a new channel of information transfer across generations. Biological and cultural evolution can intertwine in a coevolutionary spiral.

0.2 Biological steps to language-readiness -- preadaptations

A preadaptation is a change in a species which is not itself adaptive (i.e. is selectively neutral) but which paves the way for subsequent adaptive changes. Though speech is clearly adaptive, bipedalism is not itself an adaptation for speech; it is a preadaptation.

Preadaptations for language involved the following capacities or dispositions:

1. A pre-phonetic capacity to perform speech sounds or manual gestures.
2. A pre-syntactic capacity to organize longer sequences of sounds or gestures.
3. Pre-semantic capacities:
    1. to form basic concepts,
    2. to construct more complex concepts (e.g. propositions),
    3. to carry out mental calculations over complex concepts.
4. Pre-pragmatic capacities:
    1. to infer what mental calculations others can carry out,
    2. to act cooperatively,
    3. to attend to the same external situations as others,
    4. to accept symbolic action as a surrogate for real action.
5. An elementary symbolic capacity to link sounds or gestures arbitrarily with basic concepts, such that perception of the action activates the concept, and attention to the concept may initiate the sound or gesture.

If some capacity is found in species distantly related to humans, this can indicate that it is an ancient, primitive capacity. Conversely, if only our nearest relatives, the apes, possess some capacity, we can conclude that it is a more recent evolutionary development

The more ways a species is plastic in its behaviour, the more complex are the cultural traditions, including languages, that can emerge. Combined with plasticity, voluntary control adds more complexity, and unpredictability, to patterns of behaviour. Much of the difference between humans and other species can be attributed to greatly increased plasticity and voluntary control of these preadaptive capacities.

0.2.1 Pre-phonetic capacity.

Chimpanzees have little voluntary breath control. On the other hand, they have good voluntary control over their manual gestures. A preadaptation that was necessary for the emergence of modern spoken language was the extension of voluntary control from the hands to the vocal tract.

Learning controlled actions by observation entails an ability to imitate. Imitation involves an impressive `translation' of sensory impressions into motor commands. A capacity for imitation is found in a perplexing range of species. A neural basis of imitation has been found in monkeys, in the form of `mirror neurons', which fire both when an animal is carrying out a certain action, such as grasping, and when it observes that same action carried out by another animal. Our ape cousins have no trouble in recognizing different spoken human words. The capacity to discriminate the kinds of sounds that constitute speech evidently preceded the arrival of speech itself.

0.2.2 Pre-syntactic capacity.

Syntax involves the stringing together of independent subunits into a longer signal. In phonological syntax, the units, like the letters in a written word, have no independent meaning. In lexical syntax, the units, such as the words in an English sentence, have meanings which contribute to the overall meaning of the whole signal. Many bird species can learn songs with phonological syntax. . Our closest relatives, the apes, do produce long calls composed of subunits. The long calls of gibbons are markers of individual identity, for advertising or defending territory. The subunit notes, used in isolation, out of the context of long calls, are used in connection with territorial aggression, and it is not clear whether the meanings of these notes can be composed by any plausible operation to yield the identity-denoting meaning of the whole signal. Despite this variability, rules govern the internal structure of songs. Male gibbons employ a discrete number of notes to construct songs. Songs are not formed through a random assortment of notes. The use of note types varies as a function of position, and transitions between note types are nonrandom. Little is known about the ability of apes to learn hierarchically structured behaviours, although all researchers seem to expect apes to be less proficient at it than humans.

0.2.3 Pre-semantic capacities

Basic concept formation. ``Perceptual categorisation and the retention of inner descriptions of objects are intrinsic characteristics of brain function in many other animals apart from the anthropoid apes." Animals have the concepts that they need, adapted to their own physiology and ecological niche. What is so surprising about humans is how many concepts they have, or are capable of acquiring, and that these concepts can go well beyond the range of what is immediately useful.

Complex concept formation.

While apes may perhaps not be capable of storing such complex structures as humans, it seems certain that they have mental representations in predicate-argument form. Simply attending to an object is analogous to assigning a mental variable to it, which functions as the argument of any predicate expressing a judgement made by the animal. It seems certain that all species closely related to humans, and many species more distantly related, have at least this representational capacity, which is a pre-semantic preadaptation for language.

Mental calculation. Humans are not the only species capable of reasoning from experienced facts to predictions about non-experienced states of affairs. Apes and monkeys perform closest to humans in problem-solving, but their inferential ability falls short of human attainment.

0.2.4 Pre-pragmatic capacities

Mind-reading and manipulation. Recent studies amply demonstrate these manipulation and mind-reading abilities in chimpanzees. People can understand the intended import of statements whose literal meanings are somehow inappropriate. To explain how we cope with such indirectness, traditional logic has to be supplemented by the Cooperative Principle (Grice, 1975), which stipulates that language users try to be helpful in specified ways. The use of language requires this basis of cooperativeness.

There are reports of cases where an animal appears to be punished for some transgression of cooperativeness (Hauser, 1996:107-109). So the basis for cooperative behaviour, and for the instilling of such behaviour in others, exists in species closely related to humans.

Joint attention. Successful use of language demands an ability to know what the speaker is talking about. A mechanism for establishing joint attention is necessary. Primates more closely related to humans are better at following human gaze than those less closely related. Spontaneous pointing has also been observed in captive common chimpanzees. It thus appears that animals close to humans possess much of the cognitive apparatus for establishing joint attention, which is the basis of reference in language.

Ritualized action. Short greetings such as *Hello!* and *Hi!* are just act-performing words; they don't describe anything, and they can't be said to be true or false. We can find exactly such act-performing signals in certain ritualized actions of animals.

This process of dissociation between the form of the signal and its meaning can be seen as the basis of the capacity to form arbitrary asociations between signals and their meanings, discussed in the next section.

0.2.5 Elementary symbolic capacity. In the wild, there are many animals with limited repertoires of calls indicating the affective state of the animal. There is no evidence that such calls are learned to any significant degree. Thus no animal calls, as made in the wild, can, as yet, be taken as showing an ability to learn an arbitrary mapping from signal to message. Trained animals, on the other hand, especially apes, have been shown to be capable of acquiring arbitrary mappings between concepts and signals. The acquired vocabularies of trained apes are comparable to those of four-year old children, with hundreds of learned items. An ape can make a mental link between an abstract symbol and some object or action, but the circumstances of wild life never nurture this ability, and it remains undeveloped.

The earliest use of arbitrary symbols in our species was perhaps to indicate personal identity. In gibbons, territorial calls also have features which can indicate sex, rank and (un)mated condition. Chimpanzees apparently cannot suppress an instinctive response to concrete stimuli in favour of response to symbols. With few exceptions, even trained apes only indulge in symbolic behaviour to satisfy immediate desires.

Note that the levels of linguistic structure where language interfaces with the outside world, namely phonetics, semantics and pragmatics, were (apart from motor control of speech) in all likelihood relatively closer to modern human abilities than the `core' levels of linguistic structure, namely phonology and morphosyntax.

The crucial last biological step towards modern human language capacity was the arrival of a brain capable of acquiring a much more complex mapping between signals and conceptual representations, giving rise to the possibility of the signals and the conceptual representations themselves growing in complexity.

0.3 Cultural evolution of languages

0.3.1 The Two-Phase Nature of Language Transmission

Modern linguistics, preoccupied with synchronic competence, has yet to realize the potential for explaining both linguistic phenomena and linguistic noumena in terms of a cyclic relationship between the two, spiralling through time. The proportions of the innate cognitive contribution and the contribution due to empirically available patterns in the stimuli remain to be discovered. Methodologically, it is much harder to study performance data systematically.

0.3.2 Grammaticalization

At the heart of the grammaticalization theory is the idea that syntactic organization, and the overt markers associated with it, emerges from nonsyntactic, principally lexical and discourse, organization. The mechanism of this emergence is the spiralling interaction of the two phases of a language's existence, I-Language and E-Language. Through frequent use of a particular word, that word acquires a specialized grammatical role that it did not have before. And in some cases this new function of the word is the first instance of this function being fulfilled at all, in the language concerned.

Grammaticalization theory has largely been pursued by scholars concerned with relatively recent changes in languages. Nevertheless, a recurrent central theme in grammaticalization studies is unidirectionality. The general trend of grammaticalization processes is all in one direction.It follows that the general nature of languages must have also changed over time, as languages accumulated more and more grammaticalized forms.

`` ... on the basis of findings in grammaticalization studies, we have argued that languages in the historically non-reconstructible past may have been different -- in a systematic way -- from present-day languages. We have proposed particular sequences of the evolution of grammatical structures which enable us to reconstruct earlier stages of human language(s). ... such evolutions lead in a principled way from concrete lexical items to abstract morphosyntactic forms. [This] suggests, on the one hand, that grammatical forms such as case inflections or agreement and voice markers did not fall from heaven; rather they can be shown to be the result of gradual evolutions. Much more importantly, [this] also suggests that at the earliest conceivable stage, human language(s) might have lacked grammatical forms such as case inflections, agreement, voice markers, etc. so that there might have existed only two types of linguistic entities: one denoting thing-like time stable entities (i.e. nouns), and another one for non-time stable concepts such as events (i.e. verbs)." (Heine and Kuteva, 2002:394)

The origin of all grammatical morphemes (function words, inflections) is in lexical stems. In addition, I speculate that the earliest languages had: no proper names (but merely definite descriptions); no illocution markers (such as *please*); no subordinate clauses, or hypotaxis; no derivational morphology; less differentiation of syntactic classes (perhaps not even noun and verb); and less differentiation of Subject and Topic. All this is characteristic of (unstable) pidgins and reminiscent of Bickerton's construct `protolanguage', a crude Pidgin-like form of communication with no function words or grammatical morphemes. Still in the syntactic domain, Newmeyer (2000) has theorized that all the earliest languages were SOV, (once they had the noun/verb distinction). The earliest languages would have had, in their semantics: no metaphor; no polysemy; no abstract nouns; fewer subjective meanings (e.g.epistemic modals); less lexical differentiation (e.g. *hand/arm, saunter/stroll/amble*); fewer hyponyms and superordinate terms. Probably the earliest languages had simple vowel systems and only CV syllable structure.

0.3.3 Computer modelling of language evolution

Computer modellers of emerging language start from simulated populations with no language at all, and their simulations can lead to interesting results in which the populations have converged on coordinated communicative codes which, though still extremely simple, share noteworthy characteristics with human language.

There is a noticeable trend in recent computer simulations of language evolution away from modelling of the biological evolution of features of the language acquisition device (e.g. Hurford (1989, 1991), Batali (1994)). More recent simulations (e.g. those cited earlier in this paragraph) typically model the evolution of languages, via iterated learning. Kirby has rechristened this general class `iterative learning models (ILMs). They explore the interactions between pairs of strictly isolated factors relevant to the iterated learning model.

Massive advances in computing power make it possible to simulate the complex interactive dynamics of language learning by children and their subsequent language behaviour as adults, which in turn becomes the model for learning by the next generation of children. It is now possible to simulate not only the learning of a somewhat complex communication system by a single individual, on the basis of a corpus of presented examples of meaning-form pairs, but to embed such individual learning processes in a population of several hundred individuals (each of whose learning is also simulated) and to simulate the repetition of this population-wide process over many historical generations.

This research can be seen as a step up from the preceding paradigm of generative grammar. Early generative grammars were somewhat rigorously specified, and it was possible in some cases to check the accuracy of the predictions of the grammar. How the grammars themselves came to exist was not explained.

The recent simulation studies, while still in their infancy, can legitimately claim to embody rigorous claims about the precise psychological and social conditions in which grammars themselves evolve.

This strand of computational simulation research has the potential to clarify the essentials of the interaction between (a) the psychological capacities of language learners and (b) the historical dynamics of populations of learners giving rise to complex grammars resembling the grammars of real natural languages. In such simulations, a population of agents begins with no shared system of communication. The agents are `innately' endowed with certain competencies, typically including control of a space of possible meanings, an inventory of possible signals, and a capacity for acquiring grammars of certain specified sorts on exposure to examples of meaning-signal pairs. The simulations typically proceed with each generation learning from its predecessor, on the basis of observation of its communicative behaviour. At first, there is no coherent communicative behaviour in the simulated population. Over time, a coherent shared syntactic system emerges. The syntactic systems which have been achieved in this research paradigm are all, of course, simpler than real attested languages, but nevertheless possess many of the central traits of natural language syntactic organization, including recursivity, compositionality of meaning, asymmetric distribution of regular and irregular forms according to frequency, grammatical functional elements with no denotational meaning, grammatical markers of passive voice and of reflexivity, and elementary partitioning into phrases.

There has been less computer simulation of the evolution of phonological systems, but what exists is impressive. De Boer (2001) manages to approximate to the distribution of vowels systems in the languages of the world through a model in which individual agents exchange utterances and learn from each other. An early computational study (Lindblom, MacNeilage and Studdert-Kennedy, 1984) can be interpreted as modelling the processes by which syllables become organized into structured CV sequences of segments, where the emergent selected consonants and vowels are drawn from economical symmetrical sets, as is typical of actual languages.

Computer simulations, within the iterated learning framework, starkly reveal what Keller (1994) has called `phenomena of the third kind', and Adam Smith (1786) attributed to an `Invisible Hand'. Languages are neither natural kinds, like plants and animals, nor artefacts, deliberate creations of humans, like houses and cars. Phenomena of the third kind result from the summed independent actions of individuals, but are not intentionally constructed by any individual. Ant trails and bird flocks are phenomena of the third kind, and so, Keller persuasively argues, are languages. Simulations within the ILM framework strip the interaction between individuals down to a bare minimum which from which language-like systems can be shown to emerge. The key property of these models is that each new generation learns its language from a restricted set of exemplars produced by the preceding generation.

One of the most striking results of this work is this: in a population capable of both rote-learning and acquisition of rules generalizing over recurrent patterns in form-meaning mapping, a pressure exists toward an eventual emergent language that expresses meanings compositionally. No calculation of an individual agent's fitness is involved, nor does any consideration of the communicative efficacy of the language play a part. The convergence on `efficient' languages is essentially a mathematical outcome of the framework, analogous to the hexagonal cells of honeycombs. At least some of the regular compositional patterning we see in languages is the result, not of humans having an inbuilt bias towards learning languages of a certain type, but of the simple fact that languages are passed on from one generation to the next via a limited channel, a `bottleneck'. As Daniel Dennett has remarked (personal communication), this turns the familar `poverty of the stimulus' argument in relation to language acquisition on its head. The poverty of stimulus argument appealed to an intuition that human languages are learned from surprisingly scanty data. Work in the iterated learning framework shows that in order for regular compositional language to emerge, a bottleneck between the adult providers of exemplary data and the child learner is necessary. Interesting experiments show that in these models, overprovision of data (i.e. practically no bottleneck), results in no convergence on a regular compositional language.