The birth of rules

Jean Aitchison

1. Chunks versus rules

Language origin has finally become a topic of interest to mainstream linguists and others, and at long last, publications are beginning to proliferate (e.g., Aitchison 1996, Deacon 1997, Hurford et al. 1998, Jablonski and Aeillo 1998, Jackendoff 1999, Pinker and Bloom 1990, Trabant 1996). Now, as the twenty-first century begins, scholars are identifying some key questions. This paper will attempt to answer one of these: the chunks-versus-rules problem.

Language is not a unitary phenomenon: parts of it are memorised chunks, and parts of it are rule-governed. These two linguistic types co-exist throughout language, as seen, for example, in irregular versus regular verbs (Pinker 1999). Yet the historical relationship between the two phenomena is unclear. Do the memorised chunks lead to rules? Or do they co-exist as independent types? This is the key question that will be explored in this paper.

A parallel question used to be found in the child language literature. At one time, it was unclear whether youngsters' cries and babbles were separate from "real" language. This has now been settled: crying has been recognised as separate, prelinguistic behaviour. Babbling, on the other hand, becomes more sophisticated and develops into speech (Vihman 1996).

This paper is divided into four main sections. First, it will summarise some background assumptions on language origin. Second, it will outline some presumed early stages in the human development of language, highlighting the importance of convergence (co-evolution). Third, it will discuss the possible origin of rule-governed behaviour. Fourth, it will consider what current-day language can reveal about these early rules.

2. Background assumptions

Three important background assumptions are outlined below, concerning innately guided behaviour, Baldwinian evolution, and the "general uniformity principle". These will form a backdrop to the remainder of the paper.

2.1. Innately guided behaviour

For centuries, an age-old nature-nurture controversy has trundled on, swinging to and fro. Is language "natural", as when dogs naturally bark? Or is it "nurtured", as when dogs can be taught to beg? This debate has also been labelled a distinction between "hard-wired" and "soft-wired" behaviour, or between "instinct" and "learning".

Yet the sharp divide between these two types is now recognised as unrealistic. Lenneberg (1967) pointed out that language is controlled by maturation: children acquire language "naturally" provided that they are properly "nurtured" – exposed to adequate linguistic data at the relevant time. Gould and Marler (1987) speak of "innate guidance": instinct guides humans to pay attention to particular features, whose finer points then have to be learned, much as bees primarily pay attention to the odour, but also shape and colour of flowers. These characteristics lead them to sources of nectar, which they visit with increasing certainty as they learn to identify particular flowers. Birdsong provides a further example (Gould and Marler 1987, Marler 1998). A number of species of birds have an outline knowledge of their own songs, but have to learn the finer details. So the nature-nurture controversy has largely faded away as the division between the two types of behaviour has blurred (Aitchison 1998).

2.2. Baldwinian evolution

Around a century ago, James Mark Baldwin, an American psychologist, indicated how Darwin's theory of natural selection could be extended (Baldwin 1902). So-called Baldwinian evolution shows how new traits can emerge without the necessity of assuming (as did the discredited Frenchman Lamarck) that acquired responses to environmental challenges could be passed onto offspring directly. Behavioural flexibility and learning, Baldwin argued, can amplify and bias natural selection: a subgroup of a species can

move into a niche slightly different from that of its ancestors. Species members who could via natural variation withstand the cold, for example, would survive better during a fierce winter. Future generations might then inherit, and gradually enhance, a genetic predisposition for enduring freezing weather. From the point of view of language, those who had, say, better memory skills, might be those with greater chances of survival. These survivors might pass on an ability to retain a large vocabulary to their offspring.

2.3. General uniformity principle

The so-called "general uniformity principle" (Lass 1997: 28) that language in the past followed similar principles to language today, was clearly stated by William Dwight Whitney in the nineteenth century: "The nature and uses of speech ... cannot but have been essentially the same during all periods of its history... [T]here is no way in which the unknown past can be investigated, except by the careful study of its living present and recorded past, and the extension and application to remote conditions of laws and principles deduced by that study" (Whitney 1867: 34). The general uniformity principle ties in with the "uniform probabilities principle" (Lass 1997: 28), that the likelihood of any linguistic state of affairs has always been roughly the same as it is now. The practical upshot of these interlocking principles is that we can make deductions about the beginnings of language by looking at current-day language.

3. Early stages of language: humans versus other primates

Humans, like all primates, are social animals. Primate social behaviour has at least three clear characteristics: strong family ties, active within-group interaction, and a well-defined ranking order. It is not surprising, therefore, that a primate should have developed a communication system which promotes these. Yet humans alone have developed language, though chimps show some linguistic precursors.

The ways in which humans and chimps diverged are outlined below. The most noticeable differences are brain size, "theory of mind", voluntary vocalisation, and speech production abilities.

3.1. Brain size

The enlarged brain size of humans may have arisen as a spin-off of the geographical location of protohumans. Humans are now generally thought to have emerged from Africa. One increasingly accepted view is that future humans lived in an area of Africa which underwent a severe drought, possibly east of the Great Rift Valley (Kortlandt 1968, Coppens 1994), though the exact location is still under discussion. This led to a cascade of further developments – especially meat eating – which alongside other factors promoted a big brain: the prefrontal area in particular is enlarged, compared with the brain of chimps (Deacon 1997). The enlarged brain involved premature births, by ape standards, and consequently neoteny (extended childhood). These immature infants were kept close to their parents, whose vocalisations they imitated during their lengthy infancy. Yet meat-eating cannot have been the only factor promoting a big brain: brain size, group size, and use of deception all intercorrelate strongly. The latter two, particularly deception, provided the impetus to developing advanced communication skills, especially a "theory of mind".

3.2. Theory of mind

A so-called "theory of mind" was possibly crucial in language development. Signing chimps, it has been noted, restrict themselves mainly to asking for things they want, such as oranges, juice, tickles, and so on. They do not talk for the sake of talking. Humans have developed an ability to put themselves into another person's shoes. This ability is turning out to be multilayered and complex. Its origins may lie in the ability to deceive, since successful deceit requires an animal to see things from another's point of view: the nearer the primate species is to humans, the more efficient they appear to be at hoodwinking one another (Whiten and Byrne 1997). The theory of mind possibly led to symbolisation, which in turn led to the "naming insight", the realisation that a sequence of sounds can be a "name" for something – an ability which develops normally in infants, though which can be delayed in deaf humans (Schaller 1991).

3.3. Voluntary vocalisation

A less obvious linguistic forerunner is the ability to vocalise voluntarily, and equally importantly, to refrain from vocalising involuntarily. Many animals, including chimps, tend to vocalise when presented with an appropriate stimulus: chimps can under certain circumstances suppress vocalisations (Byrne 1994). But this is not an everyday occurrence. Gradually, vocalisation became normal and habitual. At some point, humans began to interact via vocalisation more than in any other way. This was a grooming replacement strategy, according to Dunbar (1996), when numbers in a group rose too high to make grooming feasible.

3.4. Speech production

Chimps have similar auditory abilities to humans. The big problem arises with speech production. Apes do not have the firm tongue, L-shaped vocal tract, and lowered larynx which, arguably, allowed humans to produce clear sounds. Some of the differences are due to the flatter faces of humans: the higher quality animal-based diet also reduced chewing requirements, which led to a reduction of face and jaw size (Aiello 1998). The lowered larynx is possibly due to humans' upright posture. Yet an inability to make clear sounds could not have been the only factor holding other apes back: language could have emerged in another medium, notably sign, as some people propose happened prior to spoken language (e.g., Armstrong et al. 1995, Givón 1995).

3.5. Convergence (co-evolution)

True language began, possibly, when some of the factors outlined above converged and amounted to more than the sum of the parts. As Givón (1995: 426) notes: "the socio-cultural, cognitive, communicative, behavioural and the neurological aspects of language probably evolved in parallel rather than serially. These profoundly interdependent, interactive changes thus co-evolved". The big step forward could have arisen when a clear output converged with the naming insight. The convergence of these two abilities possibly led to a naming "explosion", a desire to name dozens of objects (Figure 1).

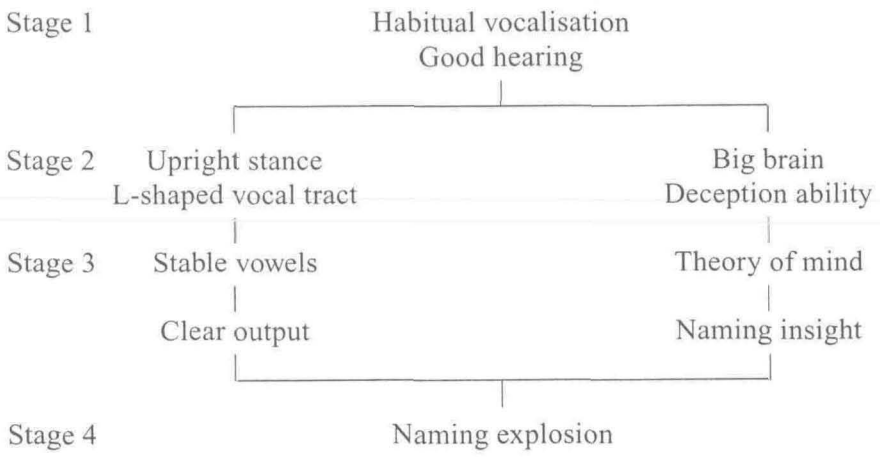

Figure 1. Convergence of abilities leading to the "naming explosion".

4. The emergence of rule-governed behaviour

A process of preferences to habits to rules (discussed further below) can be hypothesised as the basis for "real" language (Aitchison 1998). The date for the emergence of rule-governed language is still under discussion. The earliest suggested date is around 250,000 b.p., though this is most probably too early. The latest date put forward is around 50,000 b.p., though this is undoubtedly too late. Most researchers assume a date of around 100,000 b.p., though it seems increasingly likely that an earlier protolanguage existed for some time (Jackendoff 1999, Calvin and Bickerton 2000). This comprised maybe a handful of words for key needs – like *come*, *look*, and so on – with some habitual collocations.

4.1. Word combinations: inbuilt preferences

The naming explosion presumably led to word combinations. Eventually too many words existed for them to be routinely uttered singly. In some cases, early words were possibly combined fairly randomly, though were always perhaps subject to preferences, of the type seen in ape signing. The chimp Nim, for example, combined words mostly in a random way. Yet love of

eating led him always to place foods first in any sign sequence, as *banana eat Nim, banana Nim eat, banana me eat, grape eat Nim, apple me eat.* (Terrace 1979).

Some human ordering preferences are so deeply inbuilt that they can be regarded as prelinguistic, such as a preference for placing "small on large". Humans prefer to say: *The cat sat on the mat.* It would be bizarre to say: *The mat lay under the cat.* Even if the mat was the topic of conversation, the order would likely be: *On the mat, sat the cat* (Landau and Jackendoff 1993). Other human preferences can be regarded as linguistic, since they do not inevitably result in the same order, as with a liking for placing animate subjects first: an order *Horses eat hay* or *Horses hay eat* would be more likely in the languages of the world than *Eat hay horses* or *Hay eat horses*, though these are not impossible.

Another strong human preference is for action-patient closeness, though the order of action and patient is variable: an order *Monkeys eat bananas* is as likely as *Monkeys bananas eat.*

Not all languages have a preferred word order, but in a sample of 402 languages from those that do, 87 percent combined the two linguistic preferences mentioned above, with subject-object-verb (SOV) accounting for 45 percent, and SVO for 42 percent (Tomlin 1986). At the origin of language, then, it seems likely that "animate first" and "action-object closeness" predominated.

4.2. Word classes

Word order in language depends not just on habitual collocation, but on the existence of word classes. The earliest words were possibly nouns. Certainly late learners, such as deaf people who have been deprived of language, have discovered nouns before other types of word, as with the deaf Mexican Ildefonso (Schaller 1991). And the blind Helen Keller famously discovered the word for 'water' as her earliest vocabulary item (Keller 1903).

Verbs possibly arose out of nouns via re-analysis, as in the pidgin Tok Pisin, where it is easy to see how a sequence *mi singsing* 'me song-and-dance' could be interpreted either as 'I was at the song and dance festival' or 'I sang and danced'. This distinction is possibly based on a human ability to distinguish names from events. Adjectives were possibly a subdivision of verb: in some languages stative verbs and adjectives are indistinguishable, as in Miskito Coast Creole English: *if yu wud sief, yu wud ron.*

'If want to be safe, you would run' (Holm 1988: 85). Here, the word class of *sief* is unclear.

As these examples show, re-analysis is the mechanism by which numerous changes take place in languages today (Aitchison 2001). Such re-analysis would have been possible from the earliest combinations (in accordance with the uniformity principle).

5. Clues from current-day language

Following the uniformity principle, clues about language origin can be culled from a variety of present-day sources, in particular cases where strong preferences are shown in behaviour that is formulaic, but not fully rule-governed. The distinction between an unanalysed formula and a rule-governed group of words is not necessarily clear-cut. A sequence of sounds may be just that, as with a child who learned to hammer on a door shouting *obedide* ('Open the door'), though who later learned to analyse his utterance (Peters 1983). And population groups may behave differently from one another. Gleitman and Gleitman (1979) found strong differences between educated and uneducated speakers in their ability to interpret a compound such as *house-bird glass*.

In the sections below, three situations are discussed in which current day humans produce language which is quasi-formulaic, quasi-rule governed: "newsworthiness" order, sports commentating, and headlines.

5.1. Newsworthiness

Only a portion of today's languages have a fixed word order. Some American-Indian languages have a superficially free word order, though in fact work according to a "newsworthiness" principle, bringing to the front some highlighted, newsworthy item (Mithun 1992), as with Cayuga. The (translated) answers to the (translated) questions show how this works: Q. 'Who are you going with?' A. 'SAM we're going with'. Q. 'What do you want to buy?' A. 'A DRESS I'm looking for'. Q. 'How long were they there?' A. 'TWO WEEKS they were away'. Cayuga is not alone. Ngandi and Coos behave similarly. "Word order in these languages is thus based ... on the relative newsworthiness of the constituents to the discourse. An element may be newsworthy because it represents significant new information,

because it introduces a new topic, or because it points out a significant contrast" (Mithun 1992: 39). In the earliest stages of language, therefore, as in these American-Indian languages, newsworthiness may have played a vital role.

5.2. Routines: auctioneers and sports commentators

At the origin of language, humans may have had a less efficient short-term memory than humans today; they may have been less able to process speech fast, so some clues may be found in situations in which humans are pressured to speak fast. In such circumstances, they develop linguistic routines: "because it is necessary in some cases and desirable in others, to cut down the linguistic options for both speakers and hearers, speakers often resort to formulaic speech in routine contexts. Such psychological factors are essentially concerned with individual human memory and processing capacities" (Kuiper 1996: 92).

Auctioneers have developed routines and formulae for the various stages of an auction. These are repetitious utterances, though with "slots" which change, notably, the item being auctioned, and the price bid. Each stage of the auction has its own formulae, as, for example, a first bid:

How much for that?
How much for the X?
X dollars for it.
X dollars I've got for it.
X dollars for that.
I've got X dollars for them (Kuiper 1996: 61).

The closing stages also are predictable:

X gets 'em.
X buys 'em.
They go to Y.
You're the winner, Sir (Kuiper 1996: 71).

Horse racing has formulae for each section of the race, as with the start:

There they go.
They're away and racing.
They're off and racing now.
They're on their way (Kuiper 1996: 17).

The finish also has formulae:

They go to the post.
X has won it.
X got it won (Kuiper 1996: 18).

These are slotted in around other formulae, such as "loop formulae", the arrival at the end of a lap of the course, and also around individual events, such as a particular horse falling: "the horse looks as though he is walking OK so he hasn't done any serious damage to himself and the driver is quite OK" (Kuiper 1996: 14).

Children are known to cope with some aspects of language via routines (Gleason et al. 1984, Greif and Gleason 1980), as when American youngsters follow a "trick or treat" routine when they call round at neighbours' houses at Halloween (Gleason and Weintraub 1978). And successful second language learners acquire routines to help them on their way (Fillmore 1979).

In short, routines which are partly memorised, and partly analysed, in that they have "slots" into which different items can be placed, appear to be an essential part of language under pressure of time or memory. Presumably, this has always been the case.

5.3. Newspaper headlines

Newspaper headlines illustrate an intermediate stage between routines with occasional words slotted in, and true rule-governed behaviour. Noun sequences in headlines represent a headline style which became common in the 1960s in British English newspapers (Simon-Vandenbergen 1981). These noun sequences have their own incipient "rules", in that they have a high probability of appearing in a particular order, though this is not essential.

Similar sequences are found in broadsheet and tabloid newspapers. Noun sequences involving the word *murder* are documented below for a six month period in two British broadsheet newspapers: *The Times* (T) and the *Guardian* (G); and two tabloids: the *Sun* (S) and the *Daily Mirror* (M) (Aitchison, Lewis, and Naylor 2000). These noun sequences sometimes formed the whole headline, at other times they were part of a longer one:

Shotgun murder horror (M)
Street murder inquiry *reaches dead end* (G)

Within the 220 noun sequences involving *murder*, two-noun sequences predominated:

141 two-noun (64 percent): *Turk jailed for* ***wife murder*** (T)
73 three-noun (33 percent): ***Murder trial mistress*** *tells of old flames* (M)
6 four-noun (3 percent): ***WPC murder bid charge*** (S)

At first sight, these noun sequences allowed a variety of possibilities, though *murder* was preferentially near the beginning of its sequence, rather than the end. On closer examination, the words which preceded and followed *murder* were not random. They followed a limited set of patterns, which were similar across all four newspapers analysed, and across two-, three-, and four-noun sequences.

The noun immediately preceding *murder* within its sequence was most usually the victim (48 occurrences, 59 percent):

Bell murder enquiry *ends* (T)
Jury in ***child murder case*** *was misled* (G)
Bride murder trial (S)
Sex fiend wanted over ***barmaid murder*** (M)

Place preceded murder in 16 (19 percent) of the examples:

University murder case *opens* (T)
Street murder inquiry *reaches dead end* (G)
Hospital murder man's *prediction* (S)
Amazing case of the ***M50 murder*** *and the vanishing moccasins* (M)

Cause of death preceded murder in 10 cases (12 percent):

Husband guilty of ***acid-bath murder*** (T)
Barbecue smell that led to needless ***knife murder*** (G)
Hammer murder victim *cut wife out of will* (S)
Shotgun murder horror (M)

The word immediately following murder was most commonly a legal and/or abstract term (139 examples, 79 percent), as *bid*, *case*, *charge*, *enquiry*, *plot*, *quiz*, *rap*, *trial*. For example:

Murder *on tv* (T)
Teenage girls on ***murder charge*** (G)
Lorry man on ***3-girl murder rap*** (S)
Tears of ***murder trial girl*** (M)

Of the remainder, 16 percent (28 examples) referred to some involved person, such as suspect or victim:

Murder pair *jailed for life* (T)
Murder victim *looked lovely* (G)
Car murder hubby *caged* (S)
Party's off as ***murder sisters*** *get life* (M)

A number of three-noun sequences began with the noun *murder*, in which case the second word was a legal/and or abstract term, and the third usually some involved person:

Murder trial judge *praised* (T)
Murder case husband *spent wedding night with accused* (G)
Murder quiz wife *drugged* (S)
I lied to give ***murder case lover*** *an alibi* (M)

Occasionally, however, another legal/abstract term was added:

Murder case remand (G)
Grieving dad on ***murder bid rap*** (S)

The few four-noun sequences mostly involved a victim in front of the above three-noun sequences:

Boy murder charge man *in court* (G)
Carl murder quiz man *freed by cops* (S)
Wpc murder bid charge (S)

The murder headlines therefore showed a clear murder formula:

A. Victim (most likely), or place, or cause.
B. Word murder.
C. Legal or abstract term.
D. Person accused, or second legal term.

Two-word sequences were mostly A + B, or B + C; three-word sequences were mainly A + B + C, or B + C + D. The rare four-word sequences were (sometimes) A + B + C + D.

The main difference between newspapers was in vocabulary. The broadsheets preferred to use surnames, while the tabloids mainly used the first names of victims, as:

Shaughnessy murder trial (T)
Alison murder charge (S)

In line with this formal versus informal trend, the broadsheets used relatively formal vocabulary to describe humans, such as *mother*, *father*, *husband*, *child*, *friend*:

Murder case husband *spent wedding night with accused* (G)
Jury in ***child murder case*** *was misled* (G)

The tabloids, on the other hand, often referred to humans via short, informal vocabulary, such as *mum*, *dad*, *hubby*, *tot*, *pal*:

Car murder hubby *caged* (S)
Tot is silent witness to ***mum's murder*** (M)

Predictably, perhaps, the broadsheets used fairly formal legal or technical vocabulary like *charge* and *inquiry*:

Driver faces triple ***murder charge*** (T)
Murder enquiry *arrests* (G)

The tabloids used fairly informal vocabulary like *rap* and *quiz*:

Husband on ***murder rap*** (M)
Girls in ***murder quiz*** (M)

Overall, then, the similarities between newspapers were striking in that they all used the same *murder* formula, though the vocabulary slotted in to the formula differed: fairly formal in the broadsheets, shorter and more informal in the tabloids.

These *murder* headlines illustrate how groups of words become melded into a fixed order, with the most newsworthy first, an extension of the process of grammaticalisation. Grammaticalisation is most typically the demotion of a full word to a grammatical morpheme (Hopper and Traugott 1993), though is also found with larger groups of words which co-occur and coalesce: "loose, paratactic, 'pragmatic' discourse structures develop – over time – into tight, 'grammaticalised' syntactic structures" (Givón 1979: 208). This coalescence is further explored in Tabor (1993) and Tabor and Traugott (1998).

Compound nouns show some similar properties to the newspaper headlines in that co-occurrence leads to gradual coalescence and hence to a grammatical structure, as in *dinner plate*, *finger nail*, *housewife*, *rescue worker*, though the principles involved are not always clear-cut (Bauer 1998, Downing 1977, Ryder 1994).

6. Conclusion

This paper has outlined a possible scenario for the birth of rules. It argued that human language diverged from the communication systems of other primates when a number of different factors converged: speech production ability converged with the naming insight to produce a large vocabulary. An ability to distinguish people and objects from events led to different word classes and a preference for particular orders. This led to the development of

habits. Then habits became rules. It further argued that the development of preferences to habits, then habits to rules is observable today in the process of grammaticalisation.

The question posed at the beginning of the paper was on the relationship between memorised chunks and rules. The conclusion is that memorised chunks may lead to rules, though do not inevitably do so. Rules arise when words are assigned to different types (word classes) and these types are assigned a typical order. This order is likely to have arisen partly out of natural preferences and partly out of "newsworthiness". Different types of activity may have given rise to different types of "newsworthiness".