

Did our ancestors speak a holistic protolanguage?

Maggie Tallerman <maggie.tallerman@durham.ac.uk>

University of Newcastle upon Tyne, 27 April 2005

Published 2007. Did our ancestors speak a holistic protolanguage? *Lingua* 117. 579–604.

1. Introduction

Until relatively recently, most researchers saw complex language as evolving from an earlier stage involving only single words, either uttered separately, or strung randomly together in short strings to form an unordered and structureless protolanguage. Theories that take such a view are termed *SYNTHETIC* theories, since they rely on the idea that an evolving syntax takes single items and forms structure by combining them. In theories of this nature, words come first, and syntax emerges later, creating structure which utilizes these words. This position is most clearly articulated in the work of Derek Bickerton (Bickerton 1990, 1995, 1998, 2000, 2003; Calvin and Bickerton 2000) and Ray Jackendoff (1999, 2002); see also Studdert-Kennedy (1998). Bickerton envisages a stage with a vocabulary, which is formed on the basis of pre-existing cognitive concepts – the prototypes of nouns and verbs (1990: 101) – and a simple *PROTOLANGUAGE* which allows these proto-words to be juxtaposed, but which does not constitute an actual grammar. This system has little in common with primate calls:

Since language is primarily a representational system, its antecedents are to be found not in primitive forms of communication but rather in the means by which earlier [...] species represented to themselves the universe they inhabited. [...] Protoconcepts which could serve as referents for nouns and even verbs – nouns and verbs being the basic units from which other linguistic categories are derived – were in place by the time the higher primates had developed (Bickerton 1990: 100f).

This scenario assumes that (proto)nouns and verbs are categories that already have a mental representation in early hominids, and which come to be represented by words. Since at least some non-human primates also have the cognitive capacity for such concepts, this is inherently a plausible picture: ‘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’ Hurford (2003: 45). In a model such as Bickerton’s or Jackendoff’s, principles that put words together to form phrases and clauses evolve later. Work based on this model assumes a continuity between pre-human primates and early hominids, but the continuity is primarily a cognitive one, rather than one involving primate vocal communication.

Other recent work, mainly by Alison Wray and Michael Arbib, takes a very different perspective, and proposes that words emerge from longer, entirely arbitrary strings of sounds – holophrastic utterances – via a process of fractionation (Arbib 2002, 2003, 2005). This position is also assumed in much recent work in computational modelling of evolution; see for instance Kirby (2000). Wray’s work (1998, 2000, 2002a) specifically proposes that the direct antecedents of pre-language can be found in primate vocalization systems, suggesting ‘that the holistic cries and gestures of our pre-human ancestors were transformed, over a long period of time, into a phonetically expressed set of holistic message strings’ (2002a: 115). In this *HOLISTIC* approach¹, protolanguage consists of ‘a small inventory’ of indivisible utterances, used for manipulation of other individuals, and including commands and threats, greetings and

requests (Wray 1998: 52). The following quotations illustrate Wray's view of holistic protolanguage:

Protolanguage would, then, be a phonetically sophisticated set of formulaic utterances, with agreed function-specific meanings, that were a direct development from the earlier noises and gestures, and which had, like them, no internal structure (Wray 1998: 51).

[A]rbitrary phonetic representation developed not in the service of individual words but of complete, holistic utterances [...], long before words or grammar appeared (Wray 2000: 293).

In a similar vein, Arbib proposes that 'the prelanguage of early *Homo sapiens* was composed of "unitary utterances" naming events as well as a few salient actors, objects and actions, and that this preceded the discovery of *words* in the modern sense of units for compositional formation of utterances' (2003: 183). Under this approach, the presumed continuity is therefore between the noncompositional utterances found in other primates, and early hominid vocalizations (mediated by hand gestures, in Arbib's view).

For both Arbib and Wray, such holistic utterances in early hominid societies are (initially) UNANALYSED wholes, with no consistent internal regularities of form. Each utterance is initially monomorphemic, whilst representing an entire proposition. For instance, Arbib (2003: 185) proposes that 'If one starts with holophrastic utterances then statements like *baby want milk*, *brother change baby* and so on, each must be important enough for the tribe to agree on a symbol (e.g. arbitrary string of phonemes) for each one'. Wray (2000: 294) illustrates the same idea with hypothetical strings such as **tebima** 'give that to her' and **kumapi** 'share this with her'. Each string again represents a holophrastic utterance, an entire message; it is not linguistically complex. In time, according to this scenario, unanalysed (and unanalysable) material comes to be segmented into meaningful units, when, by chance, phonetically similar substrings occur in several utterances, and can be imbued with a common meaning. In the examples above, **ma** occurs in both strings, and the meaning 'her' occurs in both formulaic utterances, so **ma** comes to mean 'her'. Gradually, a set of words/morphemes emerges from the holophrases; presumably these can then be combined by whatever grammatical principles evolve subsequently.

Wray (2002a) further proposes that protolanguage comes to use both the holistic and synthetic systems, over time, so that both a holophrastic 'whole message' system and a word-based system exist simultaneously. Indeed, she suggests that both still exist side by side in modern languages. Formulaic utterances (such as *side by side*, *by and large*, *take it or leave it*, *happy birthday*, *mind your backs*, *you can't have your cake and eat it*) are seen by Wray as the modern equivalents of the early holistic messages.

The problems associated with such an account of protolanguage are, however, legion. Some have been outlined in a recent paper by Bickerton (2003), including the most obvious objection, namely that it would be almost impossible for speakers to settle on an agreed meaning when an utterance is holistic. In this paper I will address other, more specific, problems with the holistic approach. Note that I will have nothing to say about the large body

of work on formulae in fully modern language, either from the perspective of language acquisition (first or second) or language use. My main aim is a dissection of the holistic view of early protolanguage, and I examine a number of serious flaws in the arguments proposed for such a protolanguage.

The remainder of the paper is organized as follows. In Section 2, I compare the vocalization of human and non-human primates, and conclude that, contrary to Wray's proposals, primate calls do not form a good model for protolanguage. Section 3 looks at phonological and morphological properties of language, and asks whether a holistic protolanguage is plausible in light of these properties. In Section 4, I examine the properties of formulae in attested language – as these are considered by Wray to constitute the modern form of holophrastic language – in order to see whether holistic utterances have the linguistic properties necessary to form the basis for a protolanguage. This section also considers the issue of whether holistic utterances would be learnable as a protolanguage. In the final main section, Section 5, I compare the holistic system to a model for synthetic protolanguage, examining the criticisms which have been made of the latter, and consider the merits and demerits of the contrasting systems. At several points throughout, I address the question of whether a proposed holistic protolanguage would be learnable by modern humans, under the assumption that if it would not, then it could scarcely have been learnable by pre-modern hominids. My conclusion in a brief Section 6 is that the synthetic system of protolanguage is a credible and workable model for early hominid communication, whilst the holistic system is not.

2. Holistic utterances and primate calls: human vs. non-human vocalizations

One area for debate in the field of language evolution is the issue of continuity or discontinuity with the communication or cognitive systems of other animals (now rather old hat, at least as a polarized issue, as it is clear that there are both continuities and discontinuities of various kinds). Wray and Arbib both liken holistic utterances to the 'calls' of primate communication. However, their views are clearly distinct. For Arbib the similarity is purely confined to the non-compositional nature of both calls and holistic utterances: both early humans and non-human primates 'could only communicate about situations that occurred so frequently that a specific utterance was available to "name" it' (2003: 186). Arbib stresses that whereas primates have a fixed repertoire of calls, which could only be added to over many generations, early humans could (he hypothesizes) invent new utterances which could be 'acquired through learning within a community' (ibid.).

Wray, on the other hand, presumes a shared (social) function between primate calls and holistic utterances. Furthermore, she proposes that the formulaic sequences which exist in fully-fledged language still retain that function: 'when we compare the functions of formulaic sequences today with those for which holistic noise/gesture utterances appear to be used in primates, we find a strikingly close correspondence' (2000: 287).

Moreover, Wray claims (1998: 50) that in the most literal sense 'protolanguage developed *out* of the older holistic communication system' of primates (emphasis in original). As is also seen in the fuller quotation given in Section 1, she suggests that the utterances of protolanguage 'were a direct development from the earlier noises and gestures' (1998: 51). Wray thus clearly seems to regard language as HOMOLOGOUS with other primate vocalizations. However, a

number of distinct pieces of evidence add up to a forceful argument against such a view. Before I turn to this evidence, note that I am not suggesting that Wray proposes that the actual *forms themselves* found in the putative holistic protolanguage developed from primate calls, or that there is any acoustic or articulatory similarity between primate calls and utterances of protolanguage – on the contrary, she gives no indication as to where the actual holophrastic utterances might have come from. But in proposing that protolanguage literally developed out of primate calls, she is suggesting a kind of continuity which is not supported by the evidence.

First, primate vocalization is handled by different parts of the brain than human language (Myers 1976, Bradshaw and Rogers 1992, Rizzolatti and Arbib 1998, Arbib 2002, Ploog 2002), and the homologues of Wernicke's and Broca's areas exist, but do not handle vocalization in non-human primates; for instance, removal of Broca's area homologue in rhesus monkeys does not impair vocalization. In fact, as Ploog (2002) discusses, neocortical neural structures are not used for vocalization in other primates. Rizzolatti and Arbib note the following (1998: 190):

Animals' calls and human speech are undoubtedly different phenomena. Among the many aspects that differentiate them is a marked difference in the anatomical structures underlying the two behaviors. Animals calls are mediated primarily by the cingulate cortex plus some diencephalic and brain stem structures. Speech is mediated essentially by a circuit whose main nodes are the classical Broca's and Wernicke's areas, both located on the lateral cortical surface.

In Arbib's protolanguage scenario, these differences are unproblematic because speech is *not* seen as simply evolving out of the primate vocalization system (Arbib 2003: 197). The bridge is the system of mirror neurons, hypothesized to handle not only motor commands (e.g. for grasping) but also a gestural communication system, and later, speech. For Wray, though, the continuity problem persists: holistic animal calls cannot be the direct precursor to language.

Second, primate calls are largely involuntary. Byrne and Whiten (1988) offer some counterexamples, but the point is generally accepted: the majority of primate calls are certainly involuntary, whereas the absolute hallmark of language (surely including all kinds of protolanguage) is that it is under voluntary control. As Corballis (2002: 166) notes '[I]t is precisely the *lack* of voluntary control that makes primate vocal calls ill-suited to exaptation for intentional communication'. (Arbib 2003: 184 makes this point as well, in a different context.) And as Corballis also points out (*ibid.*), just as humans cannot readily produce laughter or crying on demand, so chimpanzees (and presumably other primates) cannot produce their calls on demand. Volitional control therefore has two aspects: both the fact that humans can choose to *produce* an utterance, whilst other primates cannot, and the fact that other primates cannot choose to *suppress* their utterances. Involuntary vocalizations in humans, for instance reflex swearing or grunting (often found in aphasic patients and in Tourette's Syndrome) also appear to be regulated by different sectors of the brain than ordinary linguistic utterances, and it is likely that the ancient limbic system has control over non-volitional utterances in both human and non-human primates. In this regard, Deacon notes the following:

Human laughter and sobbing are like primate calls in that they tend to be based around alternation of the presence and absence of vocal sounds, superimposed on relatively more stable mouth postures. They are bottom-up, viscerally-driven patterns. In speech, this relationship is reversed. Relatively slower tonal changes and exhalation patterns become precisely timed and subordinated to match the rapidly fluctuating articulatory movements of the mouth and tongue. Speech is thus comparatively top-down in its control (Deacon 1997: 250f).

So in this respect too, involuntary emotional vocalizations have more features in common with primate calls than do linguistic vocalizations. However, Deacon also points out that *all* human vocalizations, including laughter, are made on the outbreath², whilst the vocalizations of other primates are made on both the outbreath and the inbreath (Deacon 1997: 250). Thus there are important physiological differences between human vocalization of all kinds and that of other primates, suggesting a third ground for suspecting that language does not develop from calls.

Fourth, primate calls are for the most part genetically transmitted, whereas the utterances of any kind of protolanguage must be entirely culturally transmitted. Types of vocalization *other* than speech share more with the call systems of other primates in this respect; for instance, Burling (2000) likens laughs, cries and sobs, as well as facial expressions and bodily postures in humans to what he terms the ‘gesture-calls’ of other animals (ape calls, facial expressions, submissive postures etc.), all of which are phylogenetically determined. Burling also makes the important point (2000: 35) that unlike linguistic signals (words and signs), these innate signals in humans are characterized by the fact that they ‘grade into each other’: this applies to giggles, laughs and guffaws as well as facial expressions, and crucially, it also applies to animal gesture-calls, again suggesting that the latter are an unlikely source for protolanguage.³ It seems, then, that speech is indeed special, and thus different in origins to primate calls. In fact, Burling concludes that we should not be looking at innate calls for the origins of protolanguage at all, but rather at the set of conventionalized signs which all members of a species produce (both gestures and sounds), but which *differ* in each individual (e.g. in human infants, the ‘arms-up’ gesture, or in chimpanzee infants, the ‘nursing poke’); these have much more in common with words, since they are learned, conventional and discrete. ‘These characteristics make them a much more promising source for early language than is any part of a gesture-call system’ (Burling 2000: 37). In fact, vocalization could plausibly have been *exapted* for expressing these conventionalized signs (which later became words).

Fifth, probably the most obvious characteristic of language which is not shared by primate calls (and the one best known to linguists) is the fact that language has evolved to dissociate sound and meaning: words are built up of a finite set of meaningless speech sounds, and the meaning of the resulting unit is in no way determined by the properties of these sounds:

The dissociation of sound and meaning has no precedence in other animal vocalizations, whose signal inventories are limited and not subject to cultural modification. The dissociation is, in fact, the critical discontinuity that separates human language from other primate systems of vocal communication [...]
(Studdert-Kennedy 2000: 165).

In sum, we have (at least) five important ways in which linguistic vocalizations differ from

primate vocalizations: the former are regulated by distinct neural structures; they are under volitional control; they are physiologically dissimilar in production (and indeed perception); they are learned rather than innate; and they exhibit what Hockett (1960) termed duality of patterning. Finally, of course, Wray's idea that 'protolanguage developed *out* of the older holistic communication system' additionally disregards the critical acoustic and articulatory distinctions between primate calls of all kinds and speech.

The factors considered above go largely unmentioned in Arbib's and Wray's work. Arbib outlines the neurological differences between human and non-human primate vocalization, but neither Wray nor Arbib discusses the remaining factors; for instance, how might early hominids have evolved to *learn* their vocalizations, moving away from the innate call systems? How does duality of patterning emerge in a model that starts with phonetically-expressed holistic strings and dissects these to get words? But these are surely crucial questions in any model that bases language on primate vocalization.

Presumably, if protolanguage could be demonstrated to be a direct development from holistic animal calls, then some kind of continuity of form (holistic utterances) and function (social manipulation, in Wray's view) could be maintained. Taken together, the differences between human and non-human primate vocalizations discussed in this section add up to a strong indication that speech and language did not develop literally out of earlier primate call systems, *contra* Wray. This means that nothing is gained by proposing a holistic protolanguage: if language is not homologous with primate calls, then all the properties of language are still to be explained.

3. Phonology, morphophonology and holistic protolanguage

Consider the possible form of vocalizations in protolanguage. Proponents of the holistic approach assume that the proposed holistic utterances were long enough to be segmented into recognizable, smaller strings. But there is no inherent reason to believe that this was so; there are other logical alternatives. For instance, the set of holistic utterances could equally plausibly comprise single (proto)syllables, all differentiated from each other (*ma*, *te*, *bo*, etc.). If, for example, there were 12 consonants in the system (assuming a rather generous complement of three voiceless and three voiced stops, three nasals and three fricatives) plus five vowels, we could have a system with 60 different CV monosyllables: would that be enough for Wray's envisaged 'small inventory of phonetic sequences' (1998: 52) which the protolanguage is assumed to comprise? Let us assume that the answer is yes. Wray states that subject to memory limitations, 'strings of any length and any phonological structure might be used' in holistic protolanguage (2000: 294). But clearly, this cannot be literally true. The process of segmentation (Wray) or fractionation (Arbib) must have something to break down into words, but in the monosyllable scenario above, fractionation essentially has no material to work on. On the other hand, this kind of monosyllabic holistic utterance could turn into a word (which no longer has the meaning of a full, complex event) quite easily. But then the holistic approach is vacuous, since in effect we already have words. (The only difference between these utterances and words would be that Wray regards words as *referential* items, whilst holistic utterances are not referential, but manipulative or in some way affective. However, this distinction seems to me an entirely arbitrary one, since non-formulaic words are by no means all referential.) The

first problem, then, is in the assumption that the strings of sounds were long enough (and indeed, phonetically distinct enough) to yield analysable parts – the equivalent to morphemes.

However, notice from the foregoing that even before we get to the problems of analysis, the proposals suffer from a crucial defect. Both Wray and Arbib seem to assume that (vocal) holistic utterances were inevitably somehow already composed of the segments of modern languages – in other words, that a phonological inventory was already in place. Note, for instance, Arbib's 'arbitrary string of phonemes' (2003: 185) mentioned above. Similarly, Wray (1998: 51) talks of protolanguage being 'a phonetically sophisticated set of formulaic utterances', and Wray (2002a: 115) assumes 'a phonetically expressed set of holistic message strings'. The idea is that over time, this reservoir of messages develops into analysable strings of sounds, eventually fragmenting into words (or presumably morphemes) as chance similarities between the segment strings are noticed. But the prior existence of discrete segments is taken for granted in this scenario. Computer modelling studies have fallen into the same trap, as Studdert-Kennedy and Goldstein (2003) note with reference to Kirby (2000): Kirby's model (which also builds words from holistic utterances) is successful only because it has discrete segments built into it from the start:

[A] necessary condition of compositional syntax (discrete phonetic units) is included in the initial conditions: compositionality can only emerge, because 'holistic' utterances readily fractionate along the fault lines of their discrete components. [...] Where, then, do these phonetic break-points come from? [...] The standard units, consonants and vowels, will not do, because they and their descriptive features are purely linguistic *and therefore precisely what an evolutionary account must explain*. (Studdert-Kennedy and Goldstein 2003: 238, emphasis added)

Wray's account, as Studdert-Kennedy and Goldstein also note, suffers from the same shortcomings. So, the second problem is that holistic accounts have simply assumed the prior existence of modern phonetic segments. In Studdert-Kennedy's alternative account (1998, 2000, 2005; Studdert-Kennedy and Goldstein 2003) the proposal is that from protosyllables – a cycle of the opening and closing of the primate jaw – emerge the various phonological 'gestures' (constrictions formed by the lips, the velum and so on); this makes it possible to get consonants and vowels, and the lexicon grows in tandem with the emerging set of consonant and vowel segments. In other words, in terms of the three main factors at play in evolution (see Carstairs-McCarthy's introduction to this volume) the proposal is that category (i) – the genetic raw material available – provides the basis for the syllable, and that this is refined through natural selection, category (ii). This forms a plausible and very different scenario to Wray's and Arbib's, in which the segments are somehow already in existence.

Moving on to phonology, I propose that the holistic model, in which segments are in some unspecified way just 'there', displays a fundamental misunderstanding about how language works. Phonemes are essential to build morphemes. But phonemes are not a given, either in ontogeny or phylogeny. Each modern language has its own specific set, and what that consists of has to be figured out by the language learner. In the case of fully-fledged language, the infant uses numerous distinct cues to differentiate words and work out the word boundaries (e.g. prosodic structure such as stressed syllables; phonotactic constraints), and statistical

information derived from the distribution of sounds also appears to be important; see for instance Jusczyk (1997), Saffran and Wilson (2003) amongst many other references. Note, though, that the cues are *linguistic* cues – even the use of statistical information relies on the analysis of pre-existing words in the language which the infant is learning (this includes words made up purely for experimental purposes). None of this linguistic information would be available to a pre-linguistic hominid; therefore, there are no readily available cues to handle the fractionation of holophrastic utterances into words/morphemes.

For instance, if early hominids have no words to work from, how can they ever know which phonetic details are significant in the speech stream and which are not? *All* details, all phonetic distinctions would, in fact, have to be treated as significant, in the absence of any evidence to the contrary. Note now that *language* never functions in this way. All languages have low level phonetic distinctions which the speakers are completely unaware of because they are not salient in that particular system. Consider, for instance, the typical allophonic variation that occurs in English between aspirated and unaspirated stops: we find the unaspirated allophone [p] of a phoneme /p/ in *open, spy* but the aspirated allophone [p^h] in *apart, pie*, with the distinction depending entirely on the phonological environment. Native speakers are fully ignorant of these distinctions because [p] and [p^h] are not contrastive phones in English; typically, the substitution of one for the other will sound odd, at most. On the other hand, in Thai, aspirated /p^h/ and unaspirated /p/ have a contrastive distribution: they are different phonemes, and the substitution of one for the other results in a different meaning – hence the existence of such minimal pairs in Thai as shown in (1):

- (1) [paa] ‘forest’, [p^haa] ‘split’. (Cf. [baa] ‘shoulder’). (Davenport and Hannahs 2005: 23)

And even in English, aspiration may in fact distinguish between potentially ambiguous strings (Davenport and Hannahs 2005: 23): the phrases *peace talks* and *pea stalks* consist of the same segments in the same order, but the former has an aspirated [t^h] and the latter an unaspirated [t]. All this tacit knowledge is crucially part of the phonological system, and this differs from language to language. However, in the proposed holistic protolanguage, there would be no way for one hominid to know which distinctions were critical in another’s speech, and which were trivial (proto)allophonic variations. One speaker would have no way of knowing whether another person was using (to oversimplify greatly) the distinctions of English or the distinctions of Thai.

Next, imagine with Wray (e.g. 1998: 55f) that the holistic protolanguage has two strings like **pademe** and **mapatu**, meaning (for instance) *give her the apple* and *give me the nuts*; the idea is that in time, the chance occurrence of **pa** in both is perceived as the same substrings in each, and associated with the overlap in meaning, so that a ‘word’ for ‘give’ emerges. But why should we assume that the sequence **pa** is *pronounced* the same in both strings, at the phonetic level? In fact, it’s highly likely that it won’t be. If these were strings of a modern language, we might expect to find something like, for instance, [p^hademe] vs. [mabat^hu], with a voiceless aspirated stop in initial position and a voiced (unaspirated) stop in medial position (i.e. [p^ha] vs. [ba]). Such modification is part of the allophonic variation which typifies the phonologies of existing languages: variation cannot help but exist because once hominids have a vocal tract in anything approaching its modern form, then specific phonetic tendencies appear spontaneously. There are sets of phonetically natural distinctions, and fully-fledged languages carve up the

phonetic natural classes in various ways to establish their phonemic inventory. But in a protolanguage without a phonological system, the hominid can have no idea that the two (or more) instances of phonetically-distinct **pa** sequences are ‘the same’ in an abstract sense.

A referee for *Lingua* suggests that the foregoing objection ‘can be countered by arguing that only similar sounds occurring in similar [phonetic] contexts would be recognized as potential morphemes. If the repertoire of potential sounds is relatively small, such occurrences would not be extremely unlikely’. However, if the ‘repertoire of potential sounds’ is *too* small, then fractionation is not going to produce very many of the distinct morphemes/words which are presumed by proponents of the holistic system to appear out of the holophrastic utterances. In fully-fledged language, a small phonemic inventory is not a problem, as languages can employ a variety of strategies in vocabulary formation to supplement a small set of (phonologically) distinct segments, such as complexification of syllable types, the use of tone, or indeed, simply tolerating pairs or triplets of homophonous morphemes (cf. English *that*, *like* etc.). But full language has the advantage of linguistic structure – a fully modern phonology to exploit in the differentiation of vocabulary, and a fully modern syntax to provide cues to the word class of homophonous morphemes. None of these features would be available at the protolanguage stage.

Wray and Arbib both assume that sequences of segments which become morphemes are (somehow) extracted from holistic units. But the nub of the problem is this: you can’t have morphemes without phonemes (since morphemes are composed of phonemes) and you can’t have phonemes without words, since you have to have semantic contrasts and minimal/near-minimal pairs in order to know what the phonemes are. Therefore, morphemes can never be extracted from holistic utterances. What this means is that a phonemic inventory can be built out of the speech stream (as it is for modern infants) but a set of words *cannot* be built out of a set of random segments, because nothing will tell the hominid which are the critical contrasts – in fact, there won’t *be* any critical contrasts at that stage.

Paradoxically, furthermore, Wray’s system appears to assume not only that contrastive segments already exist, but also (since those segments can build words) that the capacity for a phonological system itself *does* already exist, but crucially, that it would somehow have been stagnant or inactive during the period in which analysis of holistic strings occurs. To understand this problem, consider Wray’s claims concerning the kind of time frame it would take for the fractionation to occur. She assumes that holistic protolanguage was stable for a long period in the *Homo erectus* era: she envisages a period ‘when the holistic system operated alone’ (2002a: 122) of up to a million years (i.e. in the *erectus* period). Presumably, then, the proposed fractionation could have taken at least hundreds of thousands of years, operating on a ‘word-by-word’ basis:

[T]here would be no onus on any individual to fully ‘crack’ the code, that is, to segment everything and come up with a complete lexicon, morphological system and grammar. Rather, these could emerge gradually over many years, through the passing down of an accumulated body of knowledge, augmented by each generation of segmenters [...], until, in the end, some sort of form, meaning, function and distribution had been assigned to every segment (Wray 2000: 298)⁴.

But there is every reason to assume that (at the very latest) once the vocal tract had developed into the form that it has in modern humans, then similar phonological principles apply, including all the principles of phonological attrition. This means that the effects of sandhi, assimilation of place and manner of articulation and of voicing etc. must all have occurred in the speech of our hominid ancestors just as they do in our languages, along with syncope, apocope (i.e. loss of sounds), metathesis, phonological splits and mergers (between segments), and so on. We can assume with confidence that such effects existed because all known phonological systems exhibit them, and they result in the phonological changes and sets of morphophonological alternations that (in one form or another) characterize fully-fledged language. Although not all kinds of sound change occur in every language, *all* languages display some of these processes: they follow ineluctably from the natural (often conflicting) tendencies imposed by the human vocal tract. So the vocalizations of a holistic protolanguage cannot escape inevitable changes of this nature, yet somehow, the assumed fractionation is presumed to be feasible on a word-by-word basis despite these shifting sands.

To illustrate the scale of the problem, consider the massive upheaval in the phonology of (say) the Romance languages (Portuguese, Spanish, French, Italian etc.) over the course of the mere thirteen hundred years or so that they took to develop from a common proto-Romance ancestor; see Green (1990: 205). The following data exemplify the point; in fact, the orthography conceals further differences in pronunciation between the modern cognates:

(2)	Italian	French	Spanish	Portuguese	
	<i>pioggia</i>	<i>pluie</i>	<i>lluvia</i>	<i>chuva</i>	‘rain’
	<i>piaga</i>	<i>plaie</i>	<i>llaga</i>	<i>chaga</i>	‘wound’ (from ‘blow/cut’)
	<i>chiamare</i>	<i>clamer</i>	<i>llamar</i>	<i>chamar</i>	‘call’ (modern Fr. ‘proclaim’)

And this is a typical situation: over a short space of time, languages change out of all recognition. In part, this is simply due to factors of category (iii): phonological systems are to a large extent self-organizing. How, then, could the fractionation have proceeded successfully over not hundreds, but hundreds of thousands of years, when the material the speakers were working on was continually slipping out of their grasp, changing the validity of any hypothesis formed by one generation and demolishing the emerging system?

The point here is that the holistic system that the speakers are attempting to analyse, generation by generation, is not plausibly a static one. So, a string that starts off as **mapatu** could soon become **mabatu**, with intervocalic voicing, or **mafatu**, with intervocalic affrication, and later perhaps **mavatu**. But this means that the sequence **pa** which occurred as the penultimate syllable in **mapatu** will *no longer* be recognized as the ‘same’ element when it occurs word-initially, as in **pademe**, where (because of the different phonetic environment) no such change has occurred. This means that the kind of analysis which Wray sees as leading to the appearance of words is going to be impossible to achieve: what might start off as two identical sequences will, in a short space of time, be differentiated by natural linguistic processes so that they would not be perceived as identical.

A referee for *Lingua* comments ‘Whether allophony and assimilation played as important a role in protolanguage as they play in modern language is also not clear. Given that most assimilations

in modern language occur in rapid, informal speech, where meaning can be inferred from context, it is less likely that comparable amounts of assimilation occurred in a holistic protolanguage'. However, this point is actually irrelevant, because given a single holophrastic utterance of any reasonable length, assimilations etc. must have occurred *within* the utterance itself, once the vocal tract had reached anything approaching its modern form – as noted above.⁵

Furthermore, over an extended period of time, as the newly-invented stems come to be fractionated out of the holistic strings, there is another problem: the inevitable development of (an embryonic) morphophonology. The evolution of morphophonology is not a problem for a synthetic system of protolanguage, since morphophonological effects inherently arise when *existing* formatives are combined into either words or phrases (Carstairs-McCarthy 2005⁶). Carstairs-McCarthy (2005: 176-7) comments as follows:

[A]t least some collocations of meaningful items seem likely to have become sufficiently frequent in proto-‘discourse’ for those items to affect one another phonologically. So there is good reason to think that proto-‘allomorphy’ existed before morphology did. That sounds paradoxical, because the term ‘allomorphy’ seems to imply the existence of morphology. But, once one appreciates that alternations of the kind that in contemporary languages are called ‘morphophonological’ are in principle independent of morphology as a component of grammar, the paradox disappears.

But this appears to be an ingrained problem for the holistic system. Once some of the ‘fractionation’ has occurred, in the holistic scenario, then there must *be* such ‘collocations of meaningful items’. But these proto-words co-exist, as Wray makes very clear, with the remaining holistic strings – under her scenario, as noted above, for what might be hundreds of millennia. Yet if the emerging stems aren’t consistently audible in a fixed form, how can the chance similarities which Wray and Arbib need in their envisaged system ever arise? How could further fractionation occur?

Modern speakers barely notice all the linguistic paraphernalia I have been discussing in this section, such as assimilation effects and developing morphophonological processes, because part of their linguistic competence is exactly to *abstract away from* the superficial, the phonetic details, and the allomorphic variation, and process just the underlying abstraction. But there can be no ‘phoneme’ or ‘morpheme’ level for Wray’s or Arbib’s protolanguage speaker to abstract towards. Instead, every tiny phonetic distinction will be significant.

There are two serious issues here. The first concerns the period in which a stable holistic system is supposed to exist – possibly for up to a million years, according to Wray (1998: 50). Stable, maybe, but *unchanging*, not possibly. This means that it’s even harder for the speakers to decide on an agreed holistic message for any given string, because any given string is constantly being eroded, assimilated, and so on. The second issue is that analysis has to occur under these same conditions of fluctuation. Again, any hypothesis in the making is quickly going to become unviable, forcing a different hypothesis to be substituted over and over. How, under these linguistically inevitable conditions, can speakers ever agree on a set of meanings in the first place, let alone fragment the holistic strings into a set of words or morphemes?

There remains one additional – and, I think, devastating – problem. The holistic scenario assumes that by chance, similar sequences will be recognized in different holophrastic utterances, and extracted from them, even though they occur in *random* positions in the utterance – for a three-syllable utterance, for instance, the sequence **pa** might occur as first or second syllable, as in the examples **pademe** and **mapatu** above, or the third syllable, as in **wilupa**. The protolanguage speaker is supposed to be able to extract the meaning ‘give’ from all three utterances. However, in attested languages with agglutinating or polysynthetic morphology – languages in which a lexical stem is likely to be surrounded by many other lexical and grammatical morphemes in each word – a morpheme *never* occurs in a random position in the string. Instead, we find a MORPHEMIC TEMPLATE, which specifies a fixed order for a verb or noun stem and all the morphemes that attach to the left and right of it: each class of morphemes has a specific position in the string. For instance, consider the following highly representative example from the Bantu language Chichewa, ‘illustrating the multiple prefixation that occurs in most Bantu languages’ (Hyman 2003: 246-247):

- (3) si-ti-dzá-ngo-mú-ményá
 NEG-SUBJ-TENSE-ASPECT- OBJ-verb stem
 ‘We will not just hit him.’ (-dzá- ‘future’, -ngo- ‘just’)

This word/sentence is ‘holistic’ in the trivial sense that the entire proposition is a single word (consisting of a stem and a set of bound affixes) but of course it is synthetic in the crucial sense that it comprises distinct morphemes, each with a fixed slot. Chichewa does not have alternative words in which the verb stem is randomly sited somewhere different in relation to (say) the object marker, or where the object and the subject markers switch places; in other words, the kind of randomness envisaged by Wray is not attested in known linguistic systems. The template is again likely to be a category (iii) – i.e. self-organizing – type of phenomenon, but we can also hypothesize that only languages which are organized in this way are learnable by humans; in other words, natural selection would not lead to languages of other types being propagated. Without such morpheme templates (which, of course, are linguistic entities, and so have no prior existence in protolanguage) how could the hominid ever extract just the right section from a random sequence?

The whole scenario would be like a modern child trying to learn an agglutinating or polysynthetic language with *no* phonological system, *no* morpheme templates, *no* minimal pairs, and (as a final insult) *no* guarantee that two adults even shared a common meaning for the same utterance! Proponents of holistic protolanguage seem to be imbuing the early hominid with a positively superhuman analytical capacity for working out linguistic systems – a capacity certainly not found in any modern human society, or indeed, in any linguistics department. And moreover, these protolanguage speakers presumably had no help from neural circuits dedicated to linguistic analysis; put another way, there was presumably no innate UG at this stage.

A reasonable alternative view of early linguistic evolution is that the vocal tract developed in response to a growing vocabulary, forming more clearly differentiated vocal gestures; see Studdert-Kennedy (1998, 2005). In this scenario, the pressure for physiological changes leading to the control of the distinct vocal organs comes from pressure to expand the lexicon. Rather than seeing segments and syllables as an evolutionary given, as Wray’s and Arbib’s models both do, in this model the sound system grows in line with the vocabulary. Holistic utterances viewed

as the direct descendants of primate calls form an unsuitable basis for fractionation into words: the material simply isn't there to fractionate, any more than it is in the calls of modern non-human primates. Instead, the phonological potential for words has to evolve, as Studdert-Kennedy's work (1998, 2000; Studdert-Kennedy and Goldstein 2003) clearly demonstrates: 'Notice that, on this account, [phonological] gestures are in no sense targets aimed at by the emerging system. They arise from random search, driven by pressures on speakers to enlarge their shared lexicons and to "sound like one another"' Studdert-Kennedy and Goldstein (2003: 249).

4. Protolanguage as a holistic system: evidence from modern formulaic language

In this section I will examine the properties of formulaic utterances in attested language, and consider the following issues relating to evolution: i) the issue of whether there is anything particularly salient or linguistically distinctive about formulae in the first place; ii) the issue of whether holistic utterances have the linguistic properties necessary to form the basis for a protolanguage; and iii) the issue of whether holistic utterances would be learnable as a protolanguage. As I will explore in more depth in the next section, one of Wray's principal claims is that a holistic protolanguage would be more successful than a synthetic system in fulfilling what she regards as the crucial role for a protolanguage: carrying out social functions, especially manipulation of others (for instance, Wray 1998: 60ff). As a prelude, this section offers a brief empirical examination of the formulaic system.

In a recent book, Wray suggests the following view of (modern) formulaic language:

If there is a standard view of what formulaic language is [...], at its heart will be something about word strings which 'break the rules'. They can break phonological rules, by displaying fewer stresses than expected and by being articulated faster and less clearly. They can break syntactic rules, by resisting pluralization, passivization, and so on, and by containing constituents which do not take on their normal grammatical function. They can break lexical rules, by containing items which are archaic or have no independent existence. And they can break semantic rules, by combining to mean something other than they ought to, and by being more idiomatic than an equivalent nonformulaic combination. (2002c: 261)

Formulaic utterances are common in modern languages, and are often opaque in their syntax and/or semantics – *the more, the merrier; he bought a pig in a poke*. But in fact they often abide entirely by existing grammatical rules, or else can be analysed in relation to ways in which they diverge from existing grammatical rules. So, an idiom or other formula in English could not have OSV word order; it couldn't consist of two conjoined strings with the conjunction at the start or end of the pair of conjuncts; it couldn't contain a regular verb that failed to agree with the subject in the 3SG present tense indicative; and so on.⁷ Formulae can, however, contain relics of older grammatical systems, such as NA word order (*court martial*) or postpositional phrases (*this view notwithstanding*) or subjunctive verbs (*If I were you*). This suggests that formulaic utterances are *parasitic* on existing syntax, emerging from earlier states of syntax via well-known processes (such as grammaticalization). In fact, the quotation from Wray above suggests

much the same: formulae are defined relative to their differences from extant grammar. Formulae themselves evolve from existing grammar, rather than providing tailor-made models for the lexicon and/or syntax.

The very lack of productivity in formulae suggests that they cannot be models for developing grammar, because historical changes typically operate on highly productive, salient forms. As just one example, consider the well-known diachronic development of the French negative marker *pas*. This developed from a noun which was an optional adjunct with the literal meaning ‘step’, used only with verbs of motion, as in *Il ne va (pas)* ‘He doesn’t go (a step)’ to occur productively (but optionally) with any verb, not just verbs of motion, where it then lost its nominal sense and was finally reanalysed as the (obligatory) negative marker. This change (N > NEG) was only possible because speakers extended the use of *pas* to other verbs; had *pas* been unproductive, it is very unlikely that processes of grammaticalization would have turned it into a new negation marker.

There is also a clear sense that most properties of formulaic utterances are not confined just to that type of expression. We can start with the syntax. For instance, in line with the quotation from Wray immediately above, it is true that not all idioms which appear to be constructed of a transitive verb plus its complement can be passivized:

- (4) *A pig in a poke was bought by John.
*The bucket was kicked by old Fred.

But it’s also the case that firstly, formulaic strings/idioms often *can* be passivized:

- (5) The students took advantage of the professor.
Advantage was taken of the professor by the students.
The professor was taken advantage of by the students.
- (6) When it started to rain John was suddenly quiet. He’d been bragging about being able to predict the weather and not needing an umbrella, but now *the wind had been taken out of his sails*. (From *Peter’s Idiom Collection*:
<http://home.t-online.de/home/toni.goeller/idioms/53.html>)
- (7) John arrived late with his sweater on back to front and *his leg was pulled* mercilessly by his workmates.

And secondly, resistance to passivization is not restricted to formulae:

- (8) *Ten pounds were cost by this book (cf. This book cost ten pounds.)
*Three kilos are weighed by this bag. (cf. This bag weighs three kilos.)

I suggest, then, that no syntactic properties are really specific to formulae. And as developments from some previous linguistic system, formulae do not readily lend themselves as models for productive change.

We turn next to the second issue, namely the question as to whether holistic utterances are

linguistically likely to form the basis for a protolanguage. Modern formulae seem to survive exactly *because* they are unanalysed, even (sometimes) unanalysable (in terms of the speaker's grammar). Why, if modern speakers don't spend any time analysing holistic strings, should early hominids have done so? Why aren't we continually breaking them down into their component parts, and thereby eliminating them from the language? Instead, what actually happens is that formulae are hugely resistant to change – they are one of the few aspects of language that persists, often for centuries, when lexicon, syntax, morphology and phonology have all changed. Hence the appearance of NA word orders (*sergeant major*, *solicitor general*), of postpositions such as *notwithstanding* and *ago*, or *over* in *the whole world over*, as well as archaic lexical items (*poke* for *bag*). The question is, then, why were holistic strings not equally resistant to decomposition in protolanguage?

Note also that there is a crucial difference between the properties of modern formulaic strings and the proposed holistic protolanguage.⁸ Holophrastic utterances in protolanguage are intended by Wray and Arbib to execute entire propositions – each of them denoting a predicate and its arguments – as the examples from their work noted throughout have made clear. Formulaic strings, on the other hand, are rarely propositions: note that even in idioms such as *X bought a pig in a poke*, *X kicked the bucket*, the subject argument has to be separately provided (there are some exceptions, such as *Heads will roll*, *The cat is out the bag* etc., but these are not the norm). This means that the connection Wray tries to establish between the form and function of holistic protolanguage and that of modern formulae is even harder to demonstrate.

The third issue for this section concerns the problem of learnability, if protolanguage comprised a set of holistic utterances. Modern holistic utterances are easy to remember because, even if they use some archaic lexical items, irregular semantics, and/or archaic syntax, they are essentially built of linguistic components that are *familiar* – normally morphemes or words, but if not familiar words, then a familiar syntax; the exceptions to this are few indeed (*higgledy-piggledy*, *abracadabra*, *Gesundheit*, etc.) and comprise nonsense words or foreign words, a set of which modern speakers clearly have the capacity to remember. Note, too, the extent to which non-English strings maintain – or get assimilated in order to attain – existing phonotactic constraints in the borrowing language. But the strings of a holistic protolanguage are simply random phonetic strings; presumably there is no requirement for the person “coining” these terms (see Wray 1998: 52) to even stick to an agreed phonemic inventory, since, as outlined earlier, there's no reason to think there *was* an agreed phonemic inventory, let alone any phonotactic constraints.

The point is that in protolanguage, holistic strings would relate to nothing except an entire message. There are no linguistic pegs to hang any aspect of them on. How many, then, would it be reasonable to assume that a hominid with a smaller brain – at the start of the *erectus* period, cranial capacity was about half that of modern *Homo sapiens* – could learn and recall?⁹ It is unclear how many holistic strings Wray considers are required in early hominid society. As noted, Wray (1998: 52) suggests ‘a small inventory of phonetic sequences’ and states that ‘a communication system that uses utterances only for a limited set of interactional functions need not be all that large’; but Wray (2000: 297), in a self-contradiction, mentions ‘a large inventory of arbitrary strings’.

In a recent book, Sverker Johansson (2005) also raises doubts as to the issue of learnability of a

holistic protolanguage, comparing the task of the early hominid to that of the modern child:

It is not obvious to me [...] why the segmentation process envisaged by Wray (2000) would be expected to work. A similar process is certainly present in modern-day language acquisition – children first acquire some stock phrases as unanalyzed wholes, and later figure out their internal structure – but that only works because these stock phrases *have* an internal structure, given by the grammar of the adults from whom the child acquires them. As an analogy for the origin of grammar, this is unsatisfactory. (2005: 234)

Many questions therefore remain unanswered. How many holistic utterances is it reasonable to assume that the hominid could learn over the course of a lifetime (of maybe 25 years)? Was there a critical period for learning protolanguage, as there is in the case of true language? If so, then speakers might have had a learning span of around 14 years (though of course in language, vocabulary can be learnt throughout life). Compare modern infants: by 24 months the average productive vocabulary is already 300 words (Boysson-Bardies 1999) and by age 6, it has risen to around 14,000 words (Carey 1978); a reasonable estimate of learning rate is an average of 9-10 words *a day* from 18 months onwards. Assuming that the input was a set of holistic utterances, could this feat conceivably have been matched, even approached, by the smaller-brained *erectus*, lacking any linguistic cues, no fixed phonemic inventory, and with only the vaguest idea of the intended meaning of the holistic string? I submit not. Jackendoff (2002: 242) takes a similar line: ‘If the symbols [of protolanguage] were holistic vocalizations like primate calls, even a thousand symbols would be impossible to keep distinct in perception and memory’.

Furthermore, note that whereas lexical vocabulary can be stored by pairing a *concept* with the arbitrary sound string used to denote it, holistic utterances must be stored by memorizing each complex *propositional event* and learning which unanalysable string is appropriate at each event. This task is harder, not simpler, than learning words as symbols, and therefore less suitable for an early protolanguage scenario.

Finally, words will never appear out of formulae unless the hominids using holistic protolanguage have both the necessary motor control and the neural capacity to recognize phonetic strings. But the holistic approach seems to assume that these speakers had a *greater* ability in both areas than would be needed for one-by-one words: the formulae are necessarily longer strings (otherwise they couldn’t be broken down) and the speakers need to recognize and utilize subparts of these longer strings. How could these abilities exist prior to the language faculty itself?

5. A holistic protolanguage vs. a synthetic protolanguage

Wray (1998, 2000, 2002a) outlines a number of criticisms of the synthetic model, to which I now turn in this section. I will also examine proposed features of a holistic protolanguage, and look at ways in which its proponents regard it as distinct from (and superior to) a synthetic protolanguage. I will dissect the stated merits of the holistic system to see if the claims made for it can be sustained, and, just as importantly, I will ask whether a synthetic protolanguage was linguistically viable and could have led to the emergence of full language.

5.1 Where do words come from?

Wray (1998: 47) criticizes the fact that ‘it is not clear where [...] referential words would have come from’ in the synthetic system. The distinction is that the holistic system is hypothesized not to have had referential words – what Wray (*ibid*) describes as ‘names for things and simple actions’ – because each whole string is equivalent to an entire proposition. A parallel problem, however, persists in Wray’s system, just as it must in any putative protolanguage scenario, because we have no better idea where holistic strings come from, or indeed where the various calls of other primates come from, than we have about the words of the first protolanguage(s) – whether or not these were holistic. Whatever restrictions were imposed by the evolutionary state of the vocal tract must have been complied with, but apart from this obvious fact we cannot say much more. This criticism is, then, a non-issue.

Wray does, however, have a view on how new holistic utterances might have come into being:

Introducing new ones would be approximately as difficult as it is for us to coin a new monomorphemic word for an abstract idea. In both cases, usage would be the main way in which it was learned and passed on, though the protolanguage message could also be deliberately demonstrated (being a manipulative message) [...] (Wray 2002a: 133, fn.18).

It is instructive, though, to consider the ways in which new lexical items enter the vocabulary of modern languages. The fact is that words are very rarely coined entirely *ex nihilo* (and certainly don’t appear to be ‘deliberately demonstrated’), but are typically made up of existing morphemes, combined in new ways to give new meanings, or alternatively, existing whole words are simply adapted to take on a new meaning. To illustrate, take just a small set of English words from the semantic field of computing: *software*, *hardware*, *hard drive*, *disk(ette)*, *upload*, *download*, *web(site)*, *attachment*, *internet*, *laptop*; each of these is either an existing whole word, given a new meaning, or else is a concatenation of existing (free and bound) morphemes in which the resulting word (often a compound) gets a new meaning.¹⁰ It is interesting to note that new lexical items are rarely monomorphemic, though, as Andrew Carstairs-McCarthy points out to me, complex items may indeed become monomorphemic in time (e.g. *cupboard*, *holiday*). Even an apparently new lexical item such as *blogger* in fact comprises three morphemes which have long been in existence (< *web* + *log* + *er*). Other words enter a language as loans from other languages, or as acronyms, or from names of people. But not one of these methods would have been available for the putative speakers of holistic protolanguage. We can, then, turn Wray’s question on its head, and ask where did the strings that were long enough to be broken down into words really come from?

An area of vocabulary growth about which a great deal is known, however, concerns the small function words and grammatical affixes which are so much a part of existing languages. We have a very good idea where grammatical morphemes come from in fully-fledged language: they are formed from lexical morphemes, specifically from nouns and verbs, via the bundle of processes known as GRAMMATICALIZATION; see for instance Bickerton (1990: 53ff), Heine and Kuteva (2002), Hurford (2003: 51ff). The null hypothesis is that the same processes were at work in the earliest forms of language, as these authors indeed explicitly argue. Assuming the correctness of this argument, then any account of protolanguage only needs to outline the

development of nouns and verbs, and no other syntactic categories.

This puts a rather different complexion on the kind of task that the protolanguage speaker would face. It is *not*, then, the case that we need an evolutionary scenario which explains the existence of words of all classes. Once nouns and verbs come into being, well-understood linguistic processes will do the rest. And it is reasonable to assume that (proto)nouns and verbs were indeed the earliest syntactic categories, if all other categories derive from them: ‘at the earliest conceivable stage [...] 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). Since prototypes of nouns and verbs are argued both by Bickerton and by Hurford to be existing *concepts* in the pre-hominid period, we can be fairly certain that *erectus* could handle nouns and verbs, but there is no reason to think that these hominids would yet have had the mental concepts for any other syntactic categories.

We can conclude that there is even more reason to doubt the idea of a holistic protolanguage if the proposal is that *all* words are carved out of holistic strings. Nouns and verbs more or less invent themselves, in the sense that the protoconcepts must be in existence before hominids split from the (chimpanzee) genus *Pan*; other word classes follow by grammaticalization, just as in the history of well-documented existing languages, as well as in the development of pidgins into full languages (creoles). And processes such as compounding create words with entirely new meanings out of existing words. To propose a holistic strategy involving fractionation is to ignore the *known* processes by which words come into being in language – and, I suggest, in protolanguage.

Finally, as David Willis (p.c.) has pointed out to me, the holistic strategy appears to represent a kind of DEGRAMMATICALIZATION, in the sense that it constructs lexical morphemes/words out of longer strings, and entails just the opposite of the semantic and phonological reduction that characterize processes of grammaticalization (see for instance the sketch of the development of French negation marker *pas* in Section 4 above). Although degrammaticalization is not unknown in true language (see, e.g., Newmeyer 1998: Chapter 5), it is definitely a very minor force in language change compared to grammaticalization. It would therefore be highly unexpected if it constituted the dominant method of language change at some stage in pre-history.

5.2 Where do meanings come from?

A further problem for the holistic system lies in the mechanisms proposed for establishing common meanings, when hominids are starting to analyse the holistic strings; see also Bickerton (2003). Recall that in Wray’s sketch of a putative fragment of protolanguage, we find holistic strings such as **tebima** ‘give that to her’ and **kumapi** ‘share this with her’, where a chance common element **ma** in time comes to mean *her*. A major problem in this regard is that logically, similar substrings must often occur in two (or more) utterances which do *not* share any common elements of meaning at least as many times as they occur in two utterances which *do* share semantic elements. For instance, suppose that a string **mabali** also contains the **ma** sequence, but means ‘put that rock down!’. What ensures that **ma** gets associated with ‘her’? The converse problem also arises; for instance, the meaning ‘her’ could be associated with strings that don’t contain the sequence **ma**, such as Wray’s **pubatu** ‘help her’ (2000: 294).

Repeated usage alone can't establish all and only the right 'regularities' in the proto-lexicon.

Wray (1998: 56, 2000: 297) suggests various answers to the problem of counterexamples in the system. One idea is that the form might be changed in some way to fit the perceived meaning. She suggests, for instance, that a hypercorrection might occur in a string like **pubatu**, 'help her', which ought to contain **ma** for 'her' but doesn't: thus, it could be turned into **pumatu**. But as she acknowledges, this depends on the sequence **ba** not having been successfully attributed with some meaning already. Furthermore, what if 'help her' is not **pubatu** but **fu**, in other words a simple monosyllable, as I suggested earlier? This kind of hypercorrection does not very evidently work in the case of monosyllabic strings (Section 3 above), so again the idea depends on assuming holistic strings long enough to be manipulated.

Alternatively, Wray suggests, the sequence **ma** could be added to the string **pubatu**, giving something like **mapubatu**. Whilst this is plausible, it seems to be going in the wrong direction, adding pieces rather than extracting them; taken to its logical conclusion, if the 'word' for 'help' turns out to be not **pubatu**, as originally hypothesized by the hominid, but **metiwa**, this string might have to be added as well, resulting in **mapubatumetiwa**. Although none of these processes would be problematic for speakers of full language, full language already has words and morphemes: protolanguage speakers have a different and much harder task. They have no agreed set of morphemes to work from, either in terms of form or meaning; they (presumably) have few or no innate principles or dedicated neural circuits handling linguistic analysis, as modern *Homo sapiens* must, and they have a smaller brain size and quite possibly a poorer memory. Quite simply, as noted above, these early hominids seem to be credited with quite extraordinary analytical powers in this scenario, powers which are not imputed to them under the alternative synthetic model.

Another possibility discussed by Wray (1998, 2000) to handle counterexamples is that the semantic space could be divided up, so that if **pubatu** turns out not to mean 'help her', it could be reinterpreted as 'help your mother' or 'help the older woman', for instance. Thus we could obtain nuances of meaning, for instance distinguishing 'give' from 'take' or 'present', or nuances of register (such as 'bloke', 'chap', 'guy', 'man'). The problem with an individual reinterpreting the meanings of holistic utterances is that the moment this happens, the community no longer has an agreed meaning for the string. One person's reinterpretation is not necessarily the same as another's. Faced with the same potential counterexamples, each speaker could choose a different solution: what ensures that common ground is ever (re)established?

The holistic scenario is, therefore, weakened by the existence of at least as many counterexamples as there could be pieces of confirming evidence for each putative word. And if numerous distinct strategies exist to deal with counterexamples, each with a completely different outcome, as illustrated above, then the chances of a community ever settling on an agreed form/meaning correspondence are remote indeed.

Interestingly, in computer simulations involving the modelling of a transition from holistic to compositional utterances, as described, for instance, by Kirby (2000), the problem of meaning inference has not been solved. It is the norm in such simulations for the learner agents to be *given in advance* the intended meanings attached to the noncompositional utterances¹¹; since these meanings take the form of whole predicate-argument structures, it is clear that the

‘learners’ would have an impossibly large task if they somehow had to work the meanings out. But of course, hominid speakers of a putative holistic protolanguage would have been faced with exactly this problem, as Bickerton (2003: 86) also notes. So the issue remains of how meaning could ever be determined in the case of monomorphemic but semantically-complex holistic utterances.

5.3 What use is an evolving grammar?

In this section, I will examine three interrelated problems which Wray sees as inhering in a synthetic protolanguage: i) lacking a full grammar, it lacks any principles that would assist in its processing; consequently, ii) it is too ambiguous to be communicatively useful; and iii) as a result, it fails to fulfil what she regards as the primary functions of protolanguage, namely achieving successful social interaction.

At several points, Wray asks what use an evolving grammar could possibly be. If we assume a synthetic system with individual words, does this confer any selectional advantage on its users? The following quotation illustrates her point:

[T]here is a critical level of complexity that must obtain for a creative grammar to be useful in expressing propositions. [...] [I]t is difficult to imagine what advantage a primitive, half-way grammar would have for its users, over the highly successful interactional systems of other primates (and therefore presumably the precursors of modern humans) that rely on holistic noise and gesture to express the wide range of functions necessary for [...] communal living [...]. (Wray 1998: 48).

Wray goes on to explore the implications of Bickerton’s position that protolanguage had no primitive grammar. Bickerton’s work does indeed suggest at several points that the principles of grammar observed in full language are all interdependent, and must therefore have emerged all at once; see, for instance, Bickerton (1990: Chapter 7) for arguments against gradualism. But the idea that ‘a primitive, half-way grammar’ is not very useful is surely challenged by studies of evolving systems; see, for instance, Pinker and Bloom’s remarks (1990: 712) on the ‘what good is 5 percent of an eye’ debate. Moreover, as Andrew Carstairs-McCarthy points out to me, a more sophisticated alien might well regard human grammar as hopelessly primitive, yet it nonetheless enables us to communicate far more effectively than would no grammar at all.

However, one problem here stems from Wray’s concept of a ‘primitive grammar’, and the fact that she assumes that a few simple principles of ordering actually constitute ‘grammar’. For instance, Wray (1998) can envisage that a principle such as ‘whichever word comes first is the topic’ (1998: 49) would be useful in protolanguage. Yet she then instantly dismisses this idea because ‘[t]his is grammar, so the system is not grammarless’. But this is a misapprehension, and nor is it consonant with Bickerton’s idea of what grammar is. Moreover, the recent work of Ray Jackendoff (2002: Chapter 8) demonstrates that it is entirely possible to conceive of a protolanguage with some basic, semi-formalized principles, including ordering, which would be extremely useful in disambiguating the potentially ambiguous utterances of a synthetic protolanguage, yet which would *in no way* constitute a fully-developed grammar, or in fact a ‘grammar’ of *any* kind.

Amongst other principles, Jackendoff (2002: 247ff) discusses three simple ordering patterns, Agent First, Focus Last, and Grouping which are, he suggests, ‘fossil principles’ from protolanguage. For instance, in *dog brown eat mouse*, the Agent First principle would reliably identify the dog as the agent. Grouping puts modifiers adjacent to the items they modify, so that *dog ate mouse brown* can only mean that the mouse is brown, and not the dog. Jackendoff continually stresses that all these principles are semantically based, and do not rely on any kind of syntactic knowledge; he notes that ‘Crucially, these principles correlate linear order with semantic roles. They do not require syntactic structure: the linear order of words can be determined directly in terms of phonological concatenation’ (2002: 249). This means that a synthetic protolanguage could function without the devastating ambiguity that Wray imagines (see below), just by concatenating words, if it adopted simple ordering restrictions. Indeed, given that it is well known that apes in language training experiments can spontaneously adopt ordering (Savage-Rumbaugh et al. 1998) and even parrots can be trained to pay attention to sequencing of symbols (Pepperberg 2000), it would be very surprising if our hominid ancestors did not share that same skill. If some of the same principles are later adapted or exapted for use in language itself, that too is not a surprising outcome.

All the principles that Jackendoff discusses *precede* the evolution of hierarchical phrase structure, which is the crucial development to full syntax; see Bickerton (1998: 342ff), Jackendoff (2002: Chapter 8.8). So protolanguage could have contained a very small number of these highly effective ordering and grouping principles, and these must surely have made a syntax-free (pre-syntactic) synthetic protolanguage an extremely viable system of communication, contrary to Wray’s assumptions. And since protolanguages (like languages) are culturally transmitted, any of these semi-formal patterns/principles which are adopted can be maintained from generation to generation without any innate language acquisition device. I emphasize again that the system is still ‘grammarless’, since these principles fall far short of syntax. Jackendoff (2002: 250) concludes:

Whatever the particular details of these sorts of principle that map between semantic roles and pure linear order, they sharpen communication. They are therefore a plausible step between unregulated concatenation and full syntax. In fact, unregulated concatenation need not necessarily have preceded the appearance of these principles [...].

Note that on Jackendoff’s view of an evolving protolanguage, we can simultaneously maintain two ideas that to date have seemed quite contradictory: Bickerton’s view that the emergence of syntax must be catastrophic, and the view of Pinker and Bloom (1990), Pinker and Jackendoff (2005) that grammar could have emerged gradually. As noted above, Bickerton has consistently (and cogently) argued that the criterial properties of full syntax are all interdependent (see the span of his work from 1990 to 2003: 91), which implies that one syntactic feature cannot evolve separately from all the others. But pre-grammatical principles of ordering and grouping such as those discussed above *can* emerge independently – and can evolve gradually. If each stage enhanced communication skills, then we can assume with Pinker and Bloom (1990) that such a model is adaptive – in other words, that factors of category (ii), natural selection, played a crucial role. These pre-linguistic principles are not (yet) part of grammar, but are nonetheless steps on the way to full language, and can be exapted for syntax in due course, perhaps when recursion is in place (see Hauser et al. 2002).

Principles of this kind also provide the complete answer to what Wray envisages as a synthetic system riddled with ambiguity, an issue she returns to often:

In Bickerton's model [...] the ambiguity is endemic and unavoidable (Wray 2000: 126).

[I]t is impossible to have a grammarless sequence of words that is both novel and conveys a consistently retrievable meaning (Wray 1998: 48).

Jackendoff's work, as outlined above, shows that this is simply incorrect on all counts: a few principles *can* create a very usable protolanguage, still without grammar.

This means that the main planks in Wray's argument against a synthetic system – that it would have been hopelessly ambiguous and would (thus) fail to fulfil the social functions which she sees as central – are also invalid. I now turn to the latter part of her argument, starting by examining the social functions that a holistic protolanguage performs, according to Wray. It is 'used for interpersonal manipulation and for the expression of group and personal identity' (2000: 293); she regards its essential function as revolving around social interaction, explicitly likening her position to Dunbar's (1996) grooming hypothesis. As we saw above, her view is that a synthetic protolanguage was not much use: 'Bickerton's protolanguage would be a poor vessel for the kind of subtle and complex social messages that we must assume the protolanguage speakers required for marshalling their lives within their society' (Wray 2002a: 117). So in her system it is critical that the holistic formulae are maintained in tandem with the emerging grammar as full language (with grammatical principles) was evolving: 'basic interaction and social cohesion would have been protected by the continued availability of the holistic system' (2000: 291).

First, I suggest that Wray places too much faith on a holistic system of vocalization for handling social interactions. She seems to set aside the whole panoply of primate features that our hominid ancestors must have had, just as we still have them today: such features as facial expressions, eye gaze, gestures, laughter, sobs, snarls, as well as biochemical signals such as pheromones. Additionally, it seems reasonable to assume that early hominids had paralinguistic vocalizations: shouts, curses, cries of pain and joy, etc. All of this machinery must have handled – and indeed still handles – a vast amount of our social and communicative behaviour. It has long been noted (generally by scholars outside of linguistics) that quite a small percentage of meaning is conveyed via explicit verbal communication; for instance, in his experiments on the communication of feelings and attitudes, the psychologist Albert Mehrabian discovered that in this one specific domain, only around 7 percent of meaning perceived by a listener is linguistically transmitted, with the remaining 93 percent conveyed paralinguistically (for instance by tone of voice) and – most particularly – by nonverbal cues such as gesture, posture, eye movement and so on (Mehrabian 1972). Even if the communication of other meanings or emotions does not rely quite so dramatically on the nonverbal, it is clear that nonverbal communication is just as important in the social interaction of human primates as it is for non-human primates.

The point of all this discussion is that Wray's work proposes a SOCIAL FUNCTION for a holistic protolanguage:

In both species [i.e. chimpanzee and human] [holistic utterances] are used for social interaction [...]. Given this cross-species correspondence, it [...] seem[s] reasonable to suppose that in protolanguage too, day-to-day social interaction was achieved by means of holistic utterances (Wray 2000: 289).

Wray (2000: 295) proposes that a holistic protolanguage would include demands, requests, threats, declarations of superiority/inferiority, mollifying utterances, and so on; it would not, she stresses, enable its speakers to say things like ‘This stone is heavy’. But crucially, the kind of propositions which Wray’s protolanguage is envisaged to express coincide exactly with those aspects of communication where *language* is most ineffectual and ancient primate features are most potent.

This means that hominids at the stage of protolanguage (let us assume a synthetic protolanguage supplemented by the kind of pre-grammatical principles outlined by Jackendoff 2002) need not – contra Wray – have been bereft of ways to continue their social communication. And since nonverbal features deal with societal requirements so successfully, there would be no motivation for our ancestors to either abandon them, or supplement them with ‘holistic message strings’ intended to handle the same needs. Wray regards a synthetic protolanguage as being a poor substitute for primate utterances for the communication of social messages. But the point is that it wasn’t a substitute: it wasn’t competing for the same function. As Bickerton (2003: 85) puts it:

Such things [as Wray’s *Give that to me!* etc.] are much more unambiguously expressed by behaviour already in an animal’s communicative repertoire, such as begging gestures or threat gestures. If the intended meaning [of some holistic utterance] is not apparent from that context the receiver would never be able to select, from a potentially infinite range of possible meanings, the one that the sender meant to express.

I therefore conclude that the social aspects of primate communication are *not* taken over by (proto)linguistic formulae.

Second, if the putative holistic strings did all the work Wray suggests, in terms of social manipulation, what possible motivation was there to change/analyse them at all? They ought to have been entirely sufficient. Presumably a likely answer to this from the holistic camp is that hominids still required, and still require, holistic utterances, so just added grammar on top (as the scenarios in Wray 1998, 2000, 2002a suggest), keeping the holistic strings to protect ‘basic interaction and social cohesion’ (Wray 2000: 291) as the grammar was developing. But then nothing is gained, since we still need to account for the emergence of grammar itself, whilst Wray’s characterization reduces to the claim that hominids have always had the capacity for holistic strings – hardly a controversial idea, but not one that advances knowledge about the evolution of complex (compositional) language.

6. Conclusion

I have examined claims that protolanguage consisted of initially holistic utterances, which were, over a long time frame, decomposed to form ‘words’, and I have argued that such proposals have no advantages of any kind over the synthetic model of protolanguage. Moreover, I have

presented arguments suggesting that a holistic protolanguage is problematic in a number of serious respects, including viability as a pre-linguistic system and learnability by evolving hominids. A holistic protolanguage is also largely superfluous, since the *concepts* of predicates and arguments – leading to verbs and nouns – are very likely to be primitive ones, as they are shared by modern non-human primates. Once actual nouns and verbs have evolved, then words of other classes can be taken to be derived from them, just as they are in full language, by well-understood processes of grammaticalization. The holistic approach also seems to run counter to the chief processes known to be at work in the diachronic development of language (such as grammaticalization rather than degrammaticalization), and it is unlikely that radically different linguistic machinery was in operation during language evolution.

On the other hand, the synthetic approach seems highly viable, and a synthetic protolanguage would certainly not have suffered from a fatal amount of ambiguity if supplemented by a very few simple, pre-grammatical principles, as outlined in Section 5.

The foregoing criticisms are not in any way intended as a commentary on or a critique of the proposed function of formulaic elements in fully modern language (see, for instance, Wray 2002c); the current paper has nothing to say in this regard. Rather, my aim has been to tease apart various aspects of the claim made primarily, though not exclusively, in Wray's work, namely that the ancestor of modern language was a holistic protolanguage.

Note finally that in any case, the holistic approach is about the way WORDS come into existence, though Arbib suggests that 'words in the modern sense co-evolved with syntax' (2003: 183), albeit without providing any specifics as to how this occurred. Some preliminary ideas on the (subsequent) evolution of syntax are sketched by Wray (2002: 124ff), where she briefly addresses the implications of combining holistic utterances so that 'two previously independent messages [are] juxtaposed' (2002: 124). As Andrew Carstairs-McCarthy points out to me, the resulting kind of topic-comment structure envisaged by Wray has the flavour of the 'Basic Variety' (Klein and Perdue 1997) which Jackendoff (2002: Chapter 8) also discusses extensively in the context of language evolution. In recent work, Bickerton has outlined in considerable detail a possible avenue for the subsequent development of syntax, under the synthetic approach; see especially the Appendix to Calvin and Bickerton (2000).

Acknowledgements

Earlier versions of this paper were presented at the Fifth International Conference on the Evolution of Language, held at the Max Planck Institute for Evolutionary Anthropology, Leipzig, in March/April 2004, and at the Linguistics Association of Great Britain Annual Meeting, University of Roehampton, August 2004. I am grateful to both audiences for stimulating discussion and for subsequent suggestions (particularly from April McMahon and David Willis). Thorough critiques by two people have improved the present version greatly: Andrew Carstairs-McCarthy and Alison Wray (who revealed herself to be one of the referees for *Lingua*). In addition, I would like to thank Michael Arbib, Derek Bickerton, S.J. Hannahs, Daniel Livingstone, Steve Mithen and Fritz Newmeyer for their encouragement and for many helpful comments on my more preliminary work on this topic. None of the foregoing is responsible for any remaining errors, nor should be assumed to agree with any views expressed in this paper.

References

- Arbib, Michael A. 2002. The mirror system, imitation, and the evolution of language. In Chrystopher Nehaniv and Kerstin Dautenhahn (eds.), *Imitation in animals and artifacts*. Cambridge, MA: MIT Press. 229-280.
- Arbib, Michael A. 2003. The evolving mirror system: a neural basis for language readiness. In Christiansen and Kirby (eds.). 182-200.
- Arbib, Michael A. 2005. An action-oriented neurolinguistic framework for the evolution of protolanguage. In Tallerman (ed.). 21-47.
- Bickerton, Derek. 1990. *Language and species*. Chicago: University of Chicago Press.
- Bickerton, Derek. 1995. *Language and human behavior*. Seattle: University of Washington Press.
- Bickerton, Derek. 1998. Catastrophic evolution: the case for a single step from protolanguage to full human language. In Hurford et al. (eds.). 341-358.
- Bickerton, Derek. 2000. How protolanguage became language. In Knight et al. (eds.). 264-284.
- Bickerton, Derek. 2003. Symbol and structure: a comprehensive framework for language evolution. In Christiansen and Kirby (eds.). 77-93.
- Boysson-Bardies, Benedicte de. 1999. *How language comes to children*. Cambridge, MA: MIT Press.
- Bradshaw, John and Lesley Rogers. 1992. *The evolution of lateral asymmetries, language, tool use and intellect*. New York: Academic Press.
- Burling, Robbins. 2000. Comprehension, production and conventionalisation in the origins of language. In Knight et al. (eds.), 27-39.
- Byrne, Richard and Andrew Whiten (eds.). 1988. *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Clarendon Press.
- Calvin, William H. and Derek Bickerton. 2000. *Lingua ex machina: reconciling Darwin and Chomsky with the human brain*. Cambridge, MA: MIT Press.
- Carey, Susan, 1978. The child as word learner. In Morris Halle, Joan Bresnan and George A. Miller (eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press. 264-293.
- Carstairs-McCarthy, Andrew. 2005. The evolutionary origin of morphology. In Tallerman (ed.). 166-184.
- Christiansen, Morten H. and Simon Kirby (eds.). 2003. *Language evolution*. Oxford: Oxford University Press.
- Corballis, Michael C. 2002. Did language evolve from manual gestures? In Wray (ed.). 161-179.
- Davenport, Mike and S. J. Hannahs. 2005. *Introducing phonetics and phonology*. Second edition. London: Hodder Arnold.
- Deacon, Terrence. 1997. *The symbolic species: the co-evolution of language and the human brain*. London: Penguin Books.
- Dunbar, Robin. 1996. *Grooming, gossip and the evolution of language*. London: Faber and Faber.
- Green, John N. 1990. Romance languages. In Bernard Comrie (ed.), *The world's major languages*. Oxford: Oxford University Press. 203-209.
- Hauser, Marc, Noam Chomsky and W. Tecumseh Fitch. 2002. The faculty of language: what is it, who has it and how did it evolve? *Science* 298: 1569-1579
- Heine, Bernd and Tania Kuteva. 2002. On the evolution of grammatical forms. In Wray (ed.).

- 376-397.
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203: 88-111.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hurford, James. 2000. The emergence of syntax. In Knight et al. (eds.), 219-230.
- Hurford, James. 2003. The language mosaic and its evolution. In Christiansen and Kirby (eds.), 38-57.
- Hurford, James, Michael Studdert-Kennedy and Chris Knight (eds.). 1998. *Approaches to the evolution of language: social and cognitive bases*. Cambridge: Cambridge University Press.
- Hyman, Larry. 2003. Suffix ordering in Bantu: a morphocentric approach. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2002*. Dordrecht: Kluwer. 245–281.
- Jackendoff, Ray. 1999. Possible stages in the evolution of the language faculty. *Trends in Cognitive Sciences* 3: 272-279.
- Jackendoff, Ray. 2002. *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Johansson, Sverker. 2005. *Origins of language: constraints on hypotheses*. Amsterdam/Philadelphia: John Benjamins.
- Jusczyk, Peter W. 1997. Finding and remembering words: some beginnings by English-learning infants. *Current Directions in Psychological Science* 6: 170-174.
- Kirby, Simon. 2000. Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In Knight et al (eds.). 303-323.
- Klein, Wolfgang and Clive Perdue. 1997. The Basic Variety, or: Couldn't language be much simpler? *Second Language Research* 13: 301-347.
- Knight, Chris, Michael Studdert-Kennedy and James R. Hurford (eds.). 2000. *The evolutionary emergence of language: social function and the origins of linguistic form*. Cambridge: Cambridge University Press.
- Mehrabian, Albert. 1972. *Nonverbal communication*. Chicago, IL: Aldine-Atherton.
- Myers, Ronald E. 1976. Comparative neurology of vocalization and speech: proof of a dichotomy. In Stevan Harnad, Horst D. Steklis and Jane Lancaster (eds.), *Origins and evolution of language and speech*. New York: New York Academy of Sciences. 745-757.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.
- Pepperberg, Irene M. 2000. *The Alex studies: cognitive and communicative abilities of Grey Parrots*. Cambridge, MA: MIT Press.
- Pinker, Steven and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13: 707-784.
- Pinker, Steven and Ray Jackendoff. 2005. The faculty of language: what's special about it? *Cognition* 95: 201–236.
- Ploog, Detlev. 2002. Is the neural basis of vocalisation different in non-human primates and *Homo sapiens*? In T. J. Crow (ed.), *The speciation of modern Homo sapiens*. Proceedings of The British Academy 106 (Oxford University Press). 121-135.
- Rizzolatti, Giacomo and Michael Arbib. 1998. Language within our grasp. *Trends in Neuroscience* 21.5: 188-194.
- Saffran, Jenny R. and Diana P. Wilson (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy* 4: 273-284.

- Savage-Rumbaugh, Sue, S. Shanker and Talbot J. Taylor. 1998. *Apes, language and the human mind*. New York: Oxford University Press.
- Smith, Andrew. 2005. Mutual exclusivity: communicative success despite conceptual divergence. In Tallerman (ed.) 372-388.
- Studdert-Kennedy, Michael. 1998. The particulate origins of language generativity. In Hurford et al. (eds). 202-221.
- Studdert-Kennedy, Michael. 2000. Evolutionary implications of the particulate principle: imitation and the dissociation of phonetic form from semantic function. In Knight et al. (eds.). 161-176.
- Studdert-Kennedy, Michael. 2005. How did language go discrete? In Tallerman (ed.). 48-67.
- Studdert-Kennedy, Michael and Louis Goldstein. 2003. Launching language: the gestural origin of discrete infinity. In Christiansen and Kirby (eds.). 235-254.
- Tallerman, Maggie (ed.). 2005. *Language origins: perspectives on evolution*. Oxford: Oxford University Press.
- Wray, Alison. 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18: 47-67.
- Wray, Alison. 2000. Holistic utterances in protolanguage: the link from primates to humans. In Knight et al. (eds.). 285-302.
- Wray, Alison. 2002a. Dual processing in protolanguage: performance without competence. In Wray (ed.), 2002b. 113-137.
- Wray, Alison (ed.). 2002b. *The transition to language*. Oxford: Oxford University Press.
- Wray, Alison. 2002c. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

FOOTNOTES

1. Rather unfortunately, there is no consensus on terminology in the literature. I use the term ‘synthetic’ for theories such as Bickerton’s which assume a word-based protolanguage which then puts words together to form phrases and clauses. I will use the term ‘holistic’ for theories such as Wray’s and Arbib’s, which assume a protolanguage containing initially unanalysable utterances which are then broken down into components that subsequently become words. This usage follows what I regard both as the more established terminology (e.g. Hurford 2000: 225; Bickerton 2003: 84ff) and also the more logical, since Bickerton’s model puts words together and Wray’s takes whole message strings apart. However, note that Wray herself (2000; 2002a) uses ‘analytic’ to refer to the kinds of processing Bickerton has in mind: see her fn. 1 (2002a: 113) for her comments.
2. Deacon’s comment is a valid one, though of course a small number of languages have a few sounds made with an ingressive airstream mechanism.
3. Vervet monkey alarm calls are not graded in this way, but vervets are far more distantly related to humans than are chimpanzees, and (of course) neither represents a direct ancestor to humans.
4. Note that Wray’s use of ‘segment’ is not intended in the technical sense used by linguists of ‘phone’; rather, she apparently intends something like ‘syllable’, or perhaps, more loosely, just ‘portion’. I will continue to use ‘segment’ only in the accepted linguistic sense.
5. Alison Wray, in her referee’s report for *Lingua*, also comments that I ‘argue that protolanguage users sharing the same hearth might suffer some breakdown of communication on the basis of sandhi or metathesis, even though, in [Tallerman’s] own account, speakers of Spanish somehow managed to carry on communicating while the Romance languages developed’. The point is, of course, that speakers of proto-Romance, early Spanish (etc.) all had a fully modern brain and linguistic system already, and a fully modern UG or innate predisposition toward language learning. What is more – crucially – they were already communicating with a *fully-evolved* language. Our putative language learner in the protolanguage period presumably had a (more or less) modern vocal tract, but *none of the other advantages*. Given some of the remaining problems for the holistic system outlined in this section, I do indeed consider that Wray is crediting early hominids with powers of linguistic analysis which they could not possibly have possessed.
6. Note that Carstairs-McCarthy (2005) also argues that ‘proto-“allomorphy” existed at a pre- or proto-grammatical stage like Bickerton’s protolanguage or Heine and Kuteva’s Stage X’.
7. In her referee’s commentary on an earlier version of this paper, Wray states that the claims made about formulae in the preceding sentence are not true. She lists a number of what she assumes are counterexamples to my examples of types of syntax which

are not found in formulaic expressions: a) Wray proposes *Murder she wrote* as an example of OSV word order. But this is not an instance of OSV word order, which would, of course, not be possible in English (outside of poetry), but an instance of a focalized or topicalized *fronted* object – not at all the same thing. b) She also suggests that we get ‘other formulations not permitted by the grammar’ such as ‘*on the up* (P Det P)’. But this is simply erroneous: the word class of *up* in this context is of course Noun, not Preposition, as is clear from the fact that it can be pluralized (*the ups and downs*). (Of course, *up* can also be a verb, as in *They upped the price*.) As another supposed example of formulae containing syntactic ‘formulations not permitted by the grammar’, she gives *come a cropper*, where an intransitive verb is used transitively. Again, it is a misperception that this is somehow exceptional: English grammar happily allows, for instance, what are known as cognate objects (e.g. *I dreamed a terrifying dream*) where an intransitive verb is used transitively, so there is nothing syntactically special about the idiom Wray cites. c) As a proposed counterexample to my claim that a formulaic expression couldn’t consist of two conjoined strings with the conjunction at the start or end of the pair of conjuncts, she suggests *Not only did he X, he Y’d too*. Again, this betrays a misunderstanding of the syntax: the (coordinating) conjunction in a pair of conjoined strings would be (for instance) *and*, *or*, *but*, and as far as I am aware, my original point is valid. Neither *too* nor *not only* are coordinators; see for instance Huddleston and Pullum (2002: Chapter 15) on the syntax of coordination. d) As a putative counterexample to my claim that a formula couldn’t contain a regular verb that failed to agree with the subject in the 3SG present tense indicative, she gives *Sure as eggs is eggs*. But of course, *be* is not a regular verb, so again, my original point is unaffected.

8. I am indebted to Daniel Livingstone for the argument expressed in this paragraph. As he also points out to me (p.c.) ‘Non-human holistic signals also do not map to predicate-argument meanings’ – thus suggesting a further critical distinction between primate calls and the kind of holistic protolanguage proposed by Wray and Arbib.
9. It would be an interesting experiment to see how many such strings with no familiar linguistic characteristics a modern speaker could memorize over (say) a sixth month period. Of course, the conditions faced by our hominid ancestors are impossible to replicate accurately, not least because all modern speakers are equipped with a fully modern linguistic brain.
10. The fact that most of my examples are compounds itself seems instructive. Compare Jackendoff’s proposal (2002: 249) that compounding is a linguistic fossil principle, which emerged before syntax; see Section 5 below.
11. Models which do not make the assumption of the explicit transfer of meaning are still very much in their infancy; see, for instance, Smith (2005).