linguistic system during first language acquisition, we should expect their writing to reflect that system from the start. This, too, is not the case. In semi-literates, Fairman (e.g., 2000) reports *taket* (take it), *in form* (inform), *a quaint* (acquaint) and *B four* (before). Guillaume (1927/1973) offers *semy* (c'est mis), *a bitant* (habitant), *a ses* (assez) and *dé coll*e (d'école). Thus is speech transcribed with strikingly little awareness of the grammatical or morphological components that are supposedly being freely manipulated.

All of these oddities are readily explained if humans are predisposed to treat input and output holistically where they can, and to engage in linguistic analysis only to the extent demanded by expressive need (rather than a principle of system) – *needs-only analysis* (NOA; Wray 2002a, pp. 130–32). Coupled with a parsimonious approach to pattern identification, NOA will:

a) Prevent the deconstruction of linguistic material that is no longer morphologically active, thus preserving irregularity;

b) Fence off combinations that are regular but are not observed to be subject to paradigmatic variation, and maintain them as complete units that cannot be generalised to other cases (as with the L1 acquisition of Esperanto); in so doing, protect the units from subsequent linguistic change, so they drift over time through fuzzy semi-regularity to full irregularity;

c) Support, in those who do not subsequently augment their fuzzy, half-formed linguistic system with formal training through literacy, a tolerance for underspecification and an absence of any expectation that language is *fully* composed of atomic lexical units. The bizarre spellings of semi-literates reflect a direct link between the whole meaning and its phonological form.

In addition, the fractionation of a holistic expression may often result in a "remainder" of phonological material that cannot be attributed a plausible meaning or function. Yet, because of (a) and (c), there may well never be a point when that material demands rationalisation – until the grammarian attempts to explain it in terms of a system it actually stands outside. Unless by haphazard or imposed hypercorrection, such irregular remainders may never be expunged and, although vulnerable to certain kinds of change, may persist in the long term, to the puzzlement of analysts (Wray 2002a) and frustration of adult language learners (Wray 2004).

Therefore, I contend that linguistic irregularity is a source of support for Arbib's proposal that compositionality is a choice rather than a fundamental in human language, and that its application is variable not absolute. Some aspects of what syntacticians are obliged to account for via complex rules may be no more than detritus from the process of fractionising unprincipled phonological strings.

If this is so, our challenge, before all the endangered languages disappear, is to recast our assumptions about prehistorical norms, by establishing what the "natural" balance is between compositionality and formulaicity in the absence of literacy and formal education. Many "fundamentals," such as the word, full classificatory potential, and inherent regularly of pattern, may come down to culture-centricity (Grace 2002) and the long-standing uneasy attempt to squeeze square pegs into the round holes of prevailing linguistic theory.

NOTES

1. This position easily supports Arbib's hypothesis (sect. 1.2) that there would be an extralinguistic human correlate of the primate mirror system for subcortical reflex vocalisations.

2. It was on the basis of this evidence that I first proposed a holistic protolanguage (Wray 1998; 2000; 2002b), but we avoid circularity since Arbib does not in any sense build his own story *upon* my proposal, he only cites it as an independently developed account consistent with his own.

Language evolution: Body of evidence?

Chen Yu[a] and Dana H. Ballard[b]

[a]*Department of Psychology and Cognitive Science Program, Indiana University, Bloomington, IN 47405;* [b]*Department of Computer Science, University of Rochester, Rochester, NY 14627.* **chenyu@indiana.edu dana@cs.rochester.edu http://www.indiana.edu/~dll/ http://www.cs.rochester.edu/~dana/**

Abstract: Our computational studies of infant language learning estimate the inherent difficulty of Arbib's proposal. We show that body language provides a strikingly helpful scaffold for learning language that may be necessary but not sufficient, given the absence of sophisticated language in other species. The extraordinary language abilities of *Homo sapiens* must have evolved from other pressures, such as sexual selection.

Arbib's article provides a complete framework showing how humans, but not monkeys, have language-ready brains. A centerpiece in hominid language evolution is based on the recognition and production of body movements, particularly hand movements, and their explicit representation in the brain, termed the mirror property.

How can we evaluate this proposal? One way is to take a look at infant language learning. The human infant has evolved to be language-ready, but nonetheless, examining the steps to competency in detail can shed light on the constraints that evolution had to deal with. In a manner similar to language evolution, the speaker (language teacher) and the listener (language learner) need to share the meanings of words in a language during language acquisition. A central issue in human word learning is the mapping problem – how to discover correct word-meaning pairs from multiple co-occurrences between words and things in an environment, which is termed reference uncertainty by Quine (1960). Our work in Yu et al. (2003) and Yu and Ballard (2004) shows that body movements play a crucial role in addressing the word-to-world mapping problem, and the body's momentary disposition in space can be used to infer referential intentions in speech.

By testing human subjects and comparing their performances in different learning conditions, we find that inference of speakers' intentions from their body movements, which we term embodied intentions, facilitates both word discovery and word-meaning association. In light of these empirical findings, we have developed a computational model that can identify the sound patterns of individual words from continuous speech using nonlinguistic contextual information and can employ body movements as deictic references to discover word-meaning associations. As a complementary study in language learning, we argue that one pivotal function of a language-ready brain is to utilize temporal correlations among language, perception, and action to bootstrap early word learning. Although language evolution and language acquisition are usually treated as different topics, the consistency of the findings from both Arbib's work and our work does show a strong link between body and language. Moreover, it suggests that the discoveries in language evolution and those in language acquisition can potentially provide some insightful thoughts to each other.

Language (even protolanguage) is about symbols, and those symbols must be grounded so that they can be used to refer to a class of objects, actions, or events. To tackle the evolutionary problem of the origins of language, Arbib argues that language readiness evolved as a multimodal system and supported intended communication. Our work confirms Arbib's hypothesis and shows that a language-ready brain is able to learn words by utilizing temporal synchrony between speech and referential body movements to infer referents in speech, which leads us to ask an intriguing question: How can the mirror system proposed by Arbib provide a neurological basis for a language learner to use body cues in language learning?

Our studies show quantitatively how body cues that signal intention could aid infant language learning. Such intentional body movements with accompanying visual information provide a nat-

ural learning environment for infants to facilitate linguistic processing. Audio, visual, and body movement data were collected simultaneously. The non-speech inputs of the learning system consisted of visual data, and head and hand positions in concert with eye gaze data. The possible meanings of spoken words were encoded in this nonlinguistic context, and the goal was to extract those meanings from raw sensory inputs. Our method first utilized eye and head movements as cues to estimate the speaker's focus of attention. At every attentional point in time, eye gaze was used as deictic reference (Ballard et al. 1997) to find the attentional object from all the objects in a scene, and each object was represented by a perceptual feature consisting of color, texture, and shape features. As a result, we obtained a temporal sequence of possible referents.

Next, a partitioning mechanism categorized spoken utterances represented by phoneme sequences into several meaning bins, and an expectation-maximization algorithm was employed to find the reliable associations of spoken words and their perceptually grounded meanings. Detailed descriptions of machine learning techniques can be obtained from Yu and Ballard (2004). The learning result is that this system can learn more than 85 percent of the correct word-meaning associations accurately, given that the word has been segmented. Considering that the system processes raw sensory data, and our learning method works in unsupervised mode without manually encoding any linguistic information, this level of performance is impressive.

Such results are very consistent with Arbib's proposal that these body constraints served to start language development on an evolutionary scale. However, this leaves unanswered the question of why *Homo sapiens evolved without language*. Arbib's argument seems to be that if a plausible sequence of steps is laid out, and the "height" or difficulty in transiting each step is small, then somehow evolution should have been compelled to follow this path. But our sequence of steps in the model of infant language learning also has small steps – recognize body movements, recognize intentions as communicated with body movements, recognize attentional objects in a scene, recognize the sounds that accompany these movements. These steps would be accessible for a variety of social species, and yet they were traversed only by humans.

Arbib makes special use of the hand representations, suggesting that perhaps humans had an edge in this category that provided the needed leverage. This is again very plausible, yet our studies show that you can get quite far just by hanging sounds on the end of the eye fixations and hand movements. From our point of view, any animal species that could communicate intention through body movement had the possibility of developing some kind of language. Hence, it is likely that some other constraints must be brought into play to account for the uniqueness of language in humans. Surprisingly, Arbib does not mention Miller's hypothesis that language is a product of sexual selection. Miller (2001) argues that the human brain must have been the kind of runaway process driven by sexual selection in a similar manner to Bower bird's nests and peacock's tails. Miller's arguments are extensively developed and show how *Homo sapiens* could have gotten a jump start on very similar species with very similar brain architectures.

Author's Response

The mirror system hypothesis stands but the framework is much enriched

Michael A. Arbib

Computer Science Department, Neuroscience Program and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520.
arbib@pollux.usc.edu **http://www-hbp.usc.edu/**

Abstract: Challenges for extending the mirror system hypothesis include mechanisms supporting planning, conversation, motivation, theory of mind, and prosody. Modeling remains relevant. Co-speech gestures show how manual gesture and speech intertwine, but more attention is needed to the auditory system and phonology. The holophrastic view of protolanguage is debated, along with semantics and the cultural basis of grammars. Anatomically separated regions may share an evolutionary history.

R1. Introduction

R1.1. The commentaries in perspective

The original mirror system hypothesis (MSH) states that:

> H1. The *parity requirement* for language in humans is met because Broca's area evolved atop the mirror system for grasping with its capacity to generate and recognize a set of actions.

The target article (TA) goes beyond MSH to distinguish a language-ready brain (equipping the child to learn a language) from a brain that "has" language (in the sense of, e.g., an innate "principles and parameters" universal grammar) and then to assert that:

> H2. Language readiness evolved as a multimodal manual/facial/vocal system with protosign providing the scaffolding for protospeech – these then co-evolved in an expanding spiral to provide "neural critical mass" for protolanguage

and further that:

> H3. Protolanguage was holophrastic – "protowords" were semantically more akin to phrases or sentences of modern language than words "as we know them."
>
> H4. Biological evolution gave humans a language-ready brain, but the emergence of human languages from protolanguage was a matter of history, not biology.
>
> H5. Whereas the original MSH focused on macaque F5 and Broca's area, F5 is part of a larger F5-PF-STS system in the macaque, and this "lifts" to a larger frontal-parietal-temporal language-ready system in the human brain.

Among them, H2 to H5 constitute an extended MSH. What needs stressing is that these four hypotheses are almost independent – and thus each must stand on its own. My response to the commentaries is grouped as follows:

Section R2 shows that complex imitation must be complemented by planning (R2.1) and viewed in developmental perspective (R2.2).

Section R3 generally endorses the role of the mirror sys-

https://doi.org/10.1017/S0140525X05460033 Published online by Cambridge University Press