

On Fossil Dinosaurs and Fossil Words

A startling recent development in paleontology has interested me because of its implications for paleolinguistics and for the history of science in general. Researchers headed by Mary Higby Schweitzer of North Carolina State University have claimed to have found soft tissues in dinosaur fossils 68 million years old (Schweitzer, *et al.* 2005). This finding has caused a sensation and aroused tremendous controversy, since it is “a matter of faith among scientists that soft tissue can survive at most for a few tens of thousands of years, not the 65 million since *T. rex* walked what’s now the Hell Creek Formation in Montana” (Yeoman 2006: 37). Nevertheless, Schweitzer has gained the support of some paleontologists, including the eminent dinosaur expert John R. “Jack” Horner of the Museum of the Rockies, and Jan Toporski of the Carnegie Institution (both co-authors of the *Science* article).

On the other hand, some scientists have insisted that the soft-tissue discoveries cannot be possible, and that they can be explained as outside contamination, or improper or incomplete testing of the samples (Yeoman 2006: 40).

Apart from the data and their analysis, it has been interesting to see the behavior of scientists when confronted with information that may overturn a scientific “dogma” (an oxymoron) that they have been taught to believe is immutable. Schweitzer reports that “I had one reviewer tell me that he didn’t care what the data said, he knew that what I was finding wasn’t possible. I wrote back and said, ‘Well, what data would convince you?’ And he said, ‘None’” (*Ibid*, p. 37). Horner adds, “Frequently in our field people come up with new ideas, and opponents say, ‘I just don’t believe it.’ [Schweitzer] was having a hard time publishing in journals (*Ibid*, p. 38).” This type of attitude is in fact very common in the scientific world, and it is well known that every new scientific paradigm has been met with the opposition of experts who insist that the new idea is impossible.

It is also well known that some of the greatest new hypotheses and paradigm shifts have been instigated by scholars who, in some way, are outsiders to the scientific establishment – the most famous example being a patent clerk named Einstein. It is of interest here that Schweitzer, a mother of three and a substitute teacher, came into paleontology in a rather unorthodox manner. Horner recalls that “she really wasn’t much of a scientist – which is good ... Scientists all get to thinking alike, and it’s good to bring people in from different disciplines. They ask questions very differently (*Ibid*, p. 38).”

I have been struck by the parallels this story holds with recent developments in paleolinguistics. Maverick linguists have published evidence claiming that taxonomically deep and temporally old language families, encompassing many of the universally accepted language families, can be demonstrated, and even that comparisons of these deep families show that all known human languages share a common origin, and that remnants of specific words of that Proto-Human language can still be found.¹ The rationale of these studies is well summarized by the mission statement of the Evolution of Human Languages Project, under the auspices of the Santa Fe Institute:²

Nevertheless, despite widespread scepticism and reluctance to tackle the problem

1 For a recent and sympathetic summary of this work see Ruhlen (2005).

2 See <http://ehl.santafe.edu>.

[of remote linguistic relationships], there are a number of scholars who believe that these obstacles are not insurmountable. Research has been going on over the past several decades that appears to indicate that larger genetic groupings are not only possible, but indeed quite plausible. It can be shown that most of the world's language families can be classified into roughly a dozen large groupings, or macrofamilies. Two sorts of evidence can be used for this purpose:

1) Even a superficial analysis of the vocabulary of a large number of linguistic families reveals numerous lexical similarities extending far beyond the borders of the smaller genetic units. They are frequently restricted to individual macrofamilies (such as Eurasiatic, Afroasiatic etc.), but a significant number of such matches have already been found between the macrofamilies themselves, pointing to the probability of common origin.

2) Classical historical linguistics has developed a very powerful tool - the comparative method - that allows the reconstruction of unattested language stages, so-called proto-languages. It turns out that whereas modern languages may vary significantly, protolanguages in various cases tend to be much more similar to one other. This is the case, e.g., with Indo-European, Uralic and Altaic: modern English, Finnish, and Turkish may have almost nothing in common, but their respective ancestors - Proto-Indo-European, Proto-Uralic and Proto-Altaic - appear to have many more common traits and common vocabulary. This means that the possibility exists of extending the time perspective and reconstructing even earlier stages of human language and much of this research has already been conducted.

The ultimate goal ... is to arrive at a stage when an absolute majority of the world's languages can be reduced to a minimum number of huge language macrofamilies, which in turn can be traced back to a Proto-Sapiens stage, should the databases provide sufficient evidence to support the hypothesis of monogenesis.

To many orthodox historical linguists these claims are anathema and some have vigorously opposed them. Terrence Kaufman declares that “a temporal ceiling of 7,000 to 8,000 years is inherent in the methods of comparative linguistic reconstruction,” and Paul Hopper dismisses long range linguistic comparison as “broad-based guesses.”³ R.L. Trask (1999) said “Like many linguists, I am deeply skeptical of the very possibility of identifying genetic links at the kind of time-depth [15,000-20,000 BP] envisaged here: my prejudice, based upon decades of experience, is that language change is just too rapid and too remorseless.” James Matisoff (1990) derides the work of Joseph Greenberg and other paleolinguists by coining the words “megalocomparison” (cf. megalomania) and “columbicubiculomania,”⁴ implying that this research is the result of mental pathologies!

It is clear from the above that developments in paleontology and paleolinguistics have many parallels. In both sciences we have scholars who have found evidence of what they think are very old phenomena (in the first case soft tissues in dinosaur fossils, in the other case linguistic “fossils” – the actual words spoken by early humans), and we have the old guard, who “know” – *a priori* – that such things are impossible, and “defend” the “received truth” against the upstarts. Let me elaborate on “fossil words”:

Some years ago Merritt Ruhlen and I published an article that claimed to document twenty-seven Proto-Human words and their descendants in languages all over the world (Bengtson & Ruhlen 1994). Some of the evidence we offered is shown (in greatly simplified form) in Table 1.

³ For a discussion of these remarks see Ruhlen (1994: 9-38).

⁴ “A compulsion to stick things into pigeonholes” (Matisoff 1990: 108).

Table 1: Three Global Etymologies⁵

Proto-Human	Africa	Western Eurasia	Eastern Eurasia	Oceania⁶	Americas
BUR⁷ 1 ashes, 2 dust, 3 smoke	Bongo <i>buru</i> -ku 1 Shilluk <i>bur</i> 1	Finnish <i>poro</i> ⁸ Malayalam <i>puḷu</i> -ti ⁹	Altai <i>pur</i> 1 Manchu <i>bur</i> -aki ¹⁰	Common Australian <i>*burin</i> 3 Tasmanian <i>būrana</i> 3	Uncasica <i>bura</i> 1 Lupaca <i>pur</i> -ka 1 Goajiro <i>purpura</i> 2
K'OLO¹¹ 1 hole, 2 anus, 3 buttock	!Kung ! <i>kāro</i> 1 Teda <i>kulo</i> 2	Latin <i>cūlu</i> -s 2,3 English <i>hole</i> (< <i>*kulo</i> -) Finnish <i>kolo</i> ¹² Akhwakh <i>q':oro</i> ¹³	Gondi <i>kula</i> 3 West Tibetan <i>kor</i> ¹⁴	¹⁵	North Yokuts <i>k'ol</i> -woʒ 1 Hopi <i>qōlō</i> ¹⁶ Tonkawa <i>ko-l</i> - was 3 Botocudo <i>kro</i> 1
MI(N) 1 what?, 2 who?	Kxoe <i>mā</i> ¹⁷ Hausa <i>mèè</i> , <i>mù</i> 1 Ancient Egyptian < <i>m</i> > 1 < <i>m(j)</i> > 2	Hebrew <i>mī</i> 2 Middle Breton <i>ma</i> 1 Georgian <i>ma</i> 1 Hungarian <i>mi</i> ¹⁸ Finnish <i>mi</i> , <i>mi</i> - <i>kä</i> ¹⁹ Batsbi <i>me</i> 2	Burushaski <i>men</i> 2 Ryukyuan <i>mī</i> 1 Mon <i>mu</i> 1	Arapesh <i>mane</i> 1 Common Australian <i>*minha</i> , <i>*minya</i> 1	Mandan <i>mana</i> 2 Wappo <i>may</i> 2 Guambiana <i>mu</i> 2

How, the critics ask, is it possible for the words in Table 1 to exist, in recognizable form, after what must be at least 50,000 years of linguistic evolution? (Recall the statements by Kaufman and Trask, above.) Because they “know” there is a “temporal ceiling of 7,000 to 8,000 years,” they dismiss the evidence above, and everything else that purports to represent a state of affairs earlier than the “temporal ceiling.” Just as it is “a matter of faith” among paleontologists that soft tissue cannot

5 Based on Bengtson & Ruhlen (1994), with some additions and modifications.

6 Including all the islands south and east of Asia and the Pacific; Papua-New Guinea, Australia.

7 Some have claimed that Ruhlen and I have attempted to “reconstruct” Proto-Human words. We have never made that claim, in the strict sense of phonological reconstruction, but the “phonetic glosses” such as **BUR** roughly represent what the original word may have sounded like.

8 ‘hot ashes, coarse dust’.

9 ‘dust, earth’.

10 ‘dust, sand’.

11 /k’/ represents a glottalized sound: /k/ accompanied by glottal closure.

12 ‘hole, crack’.

13 ‘burrow’.

14 ‘hollow in the ground, pit’.

15 There are several possible explanations for the gap here, for example: (1) The word **K’OLO** was lost by the ancestral language(s) of this area; (2) cognates may exist, but have not been recognized, due to phonetic or semantic changes; (3) cognates may exist, but have not been found due to lack of documentation.

16 ‘hole (in the ground)’.

17 ‘who, which’.

18 ‘what, which’.

19 ‘what, which’.

remain after millions of years, it is just as much a matter of faith among historical linguists that words cannot remain recognizable beyond 8,000 years.

Ruhlen and I, along with other paleolinguists, have argued that there is no evidence that a “temporal ceiling” exists in historical linguistics. In fact, there are several factors that, we think, make it possible for us to detect the traces of Proto-Human words that persist to modern times. These factors may be denoted as (a) localized phonetic conservatism, (b) random phonetic retention, (c) multilateral lexical recovery, and (d) reconstruction.

(a) Localized phonetic conservatism: Some languages tend to exhibit less phonetic change than others. Within Europe this can be demonstrated by French and its close taxonomic relative, Italian. Many French words have changed drastically from the original Latin form, for example Latin *aqua* ‘water’ has been reduced in French to a single vowel: *eau* /o/. Likewise, Lat. *cattus* (5 phonemes) ‘cat’ has become *chat* /ša/ (2 phonemes), *collum* ‘neck’ is now *cou* /ku/, etc. In Italian, on the other hand, the same words have undergone much less change: *acqua*, *gatto*, *collo*, respectively. In simple terms French words have a more “worn down” appearance than their Italian cognates. The modern Italian word for ‘water’, *acqua*, is little changed from Proto-Human **ʔAQ’WA** ‘water’ (Bengtson & Ruhlen 1994: pp. 327-328).

Lithuanian is often cited as a phonetically archaic language, thus Lithuanian words such as *naũjas* ‘new’, *sãpnas* ‘dream’, *ketvirtas* ‘fourth’, differ but little from their Vedic Sanskrit cognates dated some 3,500 years ago: *návyas*, *svápnas* ‘sleep’, and *caturthás*, respectively.²⁰ Indeed, Lithuanian is in some respects more archaic than Vedic. Finnish also seems to be archaic in this way, with many words remaining virtually the same since the Proto-Uralic stage of some thousands of years ago, so Proto-Uralic **silmä-* ‘eye’, **korwa-* ‘ear’, **jalka-* ‘foot’ are still *silmä*, *korva*, *jalka*, respectively, in modern Finnish. (Note that Finnish happens to figure in all three of the global etymologies cited in Table 1).

(b) Random phonetic retention: Another factor is the phenomenon of random phonetic retention, so that even in languages not particularly archaic phonetically, some words, by chance, escape radical phonetic changes. For example, English *water*, *new*, and *yoke* retain much the same form as was recorded in ancient Hittite some 3,100 – 3,600 years ago: <*watar*>, <*newa*-> and <*yuga*->, respectively. Rumanian *nepot* ‘grandson, nephew’ is still essentially the same (after ca. 8,000 years) as its reconstructed Indo-European ancestor **nepo-t-* ‘grandchild, sister’s son’.²¹

(c) Multilateral lexical recovery: When languages of the same family are compared, the chance of recovering the original vocabulary of their proto-language increases with the number of languages compared (Greenberg 1987). Thus, for example, if one is working with a large language family such as Sino-Tibetan (with more than 250 languages) there is a much better chance of recovering the original vocabulary of the proto-language than there is with a smaller family such as Uralic (some 24 languages). The same principle works on the global level, if we are testing the hypothesis that all human languages are genetically related (*i.e.*, descend from one original proto-language). For example, there is a widespread and ancient word for ‘tail’ found throughout Eurasia,

20 Though written down much later, it is thought that the Vedic texts were preserved by oral tradition from about 1,500 B.C.

21 This and other examples are cited by Bengtson & Ruhlen (1994: p. 287).

for example, Georgian *k'ud-*, Turkish *kuyruk* (from **kud-ruk*), Orok (Tungusic) *xudu*, Korean *k:ori*.²² However it seems to have been lost from the Indo-European family, except for one language: Latin *cauda*.²³ Out of ten branches of the Indo-European family²⁴ the word happened to be kept by only one.

(d) Reconstruction: The techniques of historical linguistic reconstruction – the reconstitution of older linguistic forms – allow us to recognize cognations that otherwise would escape us. Thus, for example, one might not suspect a common origin of Russian имя [ímya] ‘name’ and Hungarian *név* with the same meaning, but if we compare their respective ancestral forms, Proto-Indo-European **(e)nomen-* and Proto-Uralic **nimi-*, the similarity is easily seen.

Paleolinguists think that these four factors, and possibly others, make it possible for us to recover at least a fraction of the original Proto-Human lexicon. Historical linguists need to look at this evidence objectively and not dismiss it out of hand on the basis of what we have shown to be erroneous preconceptions.

References

- Bengtson, John D., and Merritt Ruhlen. 1994. “Global Etymologies.” In Ruhlen (1994), pp. 277-336.
- Greenberg, Joseph H. 1987. “A Generalization of Glottochronology to *n* Languages.” In *Language in the Americas*, pp. 341-344. Stanford University Press.
- Matisoff, James A. 1990. “On megalocomparison.” *Language* 66.1: 106-120.
- Ruhlen, Merritt. 1987. *A Guide to the World's Languages: Volume 1: Classification*. Stanford University Press.
- 1994. *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford University Press.
- 2005. “Taxonomy, Typology and Historical Linguistics.” In *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, ed. by J.W. Minett & W. S-Y. Wang, pp. 341-368. City University of Hong Kong Press.
- Schweitzer, M. H., J.L. Wittmeyer, J.R. Horner, J.K. Toporski. 2005. “Soft-Tissue Vessels and Cellular Preservation in *Tyrannosaurus rex*.” *Science* 307: 1952-1955.
- Trask, R.L. 1999. “Why should a language have any relatives?” In *Nostratic: Examining a Linguistic Macrofamily*, ed. by C. Renfrew & D. Nettle, pp. 157-176. Cambridge, UK: McDonald Institute for Archaeological Research.
- Yeoman, Barry. 2006. “Schweitzer’s Dangerous Discovery.” *Discover* April 2006: 37-41, 77.

22 Cf. also (with semantic shift) Uralic **kuttV* > Hungarian *hát* ‘back, backside’, etc.

23 Lat. *cauda* gave rise to It. *coda*, Rum. *coadă*, Port. *cauda*, Span. *cola* (with an unexpected shift of *d* > *l*), Fr. *queue*, etc. The long-range comparison (with Georgian, Turkish, etc.) was cited by the Russian paleolinguist Vladimir Illich-Svitych, and before that by the Italian Alfredo Trombetti.

24 Ruhlen (1987) cites Anatolian, Armenian, Tocharian, Indo-Iranian, Albanian, Greek, Italic, Celtic, Germanic, and Balto-Slavic.

