A Proto-Human Language: Fact or Fiction

Christine Schreyer

The University of Western Ontario
A Proto-Human Language: Fact or Fiction

Keywords
language, origin of language, proto-language, comparative linguistics

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
INTRODUCTION

The question of where and when anatomically modern humans emerged has long been a controversial issue in anthropology. Evidence from many different disciplines has been increasingly useful in attempting to pinpoint the origins of Homo sapiens in time and place. As it is suggested that anatomical moderns were the first humans to have the cognitive capacity for modern linguistic capabilities (Lieberman 1984), it may be plausible to use linguistic evidence in analyzing this biological evolution. This is because one of the main tenets of historical linguistics is that "languages spread as people move" (Barbujani and Pilastro 1993). In fact, Dolukhanov writes in his examination of the archaeological record that:

The advent of language in the Upper Paleolithic was a powerful factor that through the intensification of information transfer both in time (from one generation to the next) and space (between neighbouring groups) drastically accelerated social and cultural evolution. (1993).

The purpose of this paper is to determine whether or not linguistics can truly be used to uncover the origins of anatomical modern humans, or if there is a time limit at which linguistic comparisons can no longer be made. This timeframe within which linguistic comparisons can be established will also be discussed with consideration of Proto-Human, and the idea that "all the languages spoken on earth (roughly 5,000...) are descendant of a single ancestral language" (Ruhlen 1994b). This paper will describe comparative linguistics, and the many controversies which are present in this type of study as well as give examples of linguistic families which have been compiled. It will also look at genetic evidence, and the ability of this to support the linguistic data. Examples for and against the ability for comparative linguistics to compile a proto-language that encompasses all human languages will also be discussed in an attempt to determine how far these studies are viable into the past.

COMPARATIVE LINGUISTICS

Comparative linguistics, sometimes called historical linguistics, is often used in attempts to link languages that have similar cognates, phonology, and grammatical structures together. These languages which are grouped together are called language families, and are said to be linked genetically. The languages which are related are then combined into proto-languages, or languages which are "reconstructed out of the evidence that is acquired by the careful comparison of the daughter languages [or lower level languages]" (Haas 1978). These proto-languages are only estimations of original languages, however, because it is impossible to know exactly which characteristics from each daughter
language were present in the original language. Proto-languages of spoken or recorded languages are fairly accurate although estimations which move further away from the actual data are more distorted, and are therefore less demonstrable. This is one of the major problems with comparative linguistics of deep-timeframes, and is an essential fact to remember when attempting to compile an original language which may have been ancestral to all languages on earth although advocates of this theory such as Greenberg and Ruhlen do not seem to realize this. Another problem with historical linguistics is the lack of an absolute chronology.

Colin Renfrew writes that, "there seems nothing, however, in the pattern of linguistic change which would allow for the establishment of any kind of absolute chronology" (1990). He suggests that linguists take it for granted that languages change at a constant rate, and that theories based on this premise such as glottochronology are erroneous in nature. Glottochronology was a formula originally developed by Morris Swadesh in the 1960's as a means to "predict the history of the derivation from a common ancestor of a group of interrelated languages reduced to sample wordlists" (Guy 1980). It has fallen under universal criticism, however, partially as a result of "too much theorizing and little experimenting" (Guy 1980). This still leaves comparative linguistics without a model for chronology although there have been many attempts to find one. There is also the question of why languages would diverge at a constant rate when there are numerous reasons why they would not such as geographical location and isolation, population densities, and the state of relations within the speech community.

Another problem in the field of comparative linguistics is the basic assumption that comparisons "account for similarities which cannot be attributed to chance" (Hock and Joseph 1996), and are therefore the result of a common ancestor. The controversy stems from a belief that the cognates which comparative linguists use are not a product of related languages, but are due to the processes of diffusion, borrowing, and onomatopoeia. These aspects do occur, but are relatively easy to identify as chance rather than the result of genetic relations while lexical similarities which are systematic and recurring are more likely to be due to relations as are the presence of idiosyncrasies in languages (Hock and Joseph 1996). Borrowing is the use of words from one language into another, and "will generally be found in language families that are geographically adjacent...(though prehistoric adjacency might be hard to judge)" (Nettle 1999). It stands to reason then that languages that are geographically far apart will have less of a chance of having similar words unless they truly are related. Borrowing also tends to be limited to certain parts of vocabulary which are often technical in nature. In contrast similar words which are part of any basic vocabulary are likely to be the result of genetic relationships (Hock and Joseph 1996). Onomatopoeic expressions or sounds of nature also tend to be remarkably similar in languages which are not related as are "nursery words" or words "which adults assign to the early babbling of infants" (Hock and Joseph 1996).

The higher the number of languages which can be compared results in the probability of fewer chance similarities. However, using comparisons of too many languages can also be problematic as is seen in the criticisms of Joseph Greenberg's methods of mass
comparisons to create exceptionally large proto-languages from families of languages which are combined into lower level proto-languages. The mass comparison or multi-lateral comparison method establishes genetic affinities of languages through comparison of a large number of languages on a fairly limited basis (Hock and Joseph 1996). This is related to one last criticism of comparative linguistics which is that much of the work has been completed by one highly dedicated individual, and for work to be more widely accepted it is necessary for movement beyond "the great scholar" (Nettle 1999). This is partially due to the biases which individual researchers may bring to their comparisons. People with different knowledge may come up with different results even when using the same methodology (Nettle 1999).

**EXAMPLES OF PROTO-LANGUAGE**

Greenberg has used his multilateral comparison method in the midst of much criticism to suggest that the Americas are compiled of three different linguistic groups each of which coincides with a separate migration into the New World. Since traditional linguists suggest that comparisons of languages cannot be made realistically past 7,000 years they eliminate the possibility of genetic affinity between American languages and populations with Asian languages and populations as this predates the arrival of people into the New World. Greenberg's comparisons negate this notion, and support the relationship between the two groups of people. The first migration which he suggests took place approximately 11,000-12,000 years ago, and coincides with the proto-language of Amerind. Amerind is believed to be the oldest language migration into North America because it has not only reached the furthest geographical regions from which it came, but also because it "shows greater internal differentiation" (Greenberg, Turner, and Zegura 1986). The second migration is the arrival of people who speak Na-Dene, and who populated mostly the North West coast of North America, and are therefore geographically less peripheral than the third language group of Eskimo-Aleut. The arrival of the Na-Dene language group appears to be approximately 5,000 years B.P. (Krauss 1973). The third migration which is probably the most recent around 4,000 B.P. is the one which coincides with the linguistic family of Eskimo-Aleut, and whose populations live in the Northernmost part of North America. The languages and genes of this linguistic group have been shown to have relationships with Asian populations.

Amerind is the most controversial of the groupings, and this is partially due to the fact that it contains the highest amount of language families. As well it is thought to be of the greatest antiquity. There is also the controversy over the fact that Greenberg uses oral languages which have not all been extensively researched to complete his comparison. Arguments against his method include the idea of Pan-Americanism which is the belief that there is a similarity of words in North American languages simply due to similarities between Native cultures. As well, the general belief which most linguists hold is that there is simply too much to be compared, and that Greenberg can not thoroughly compare all of the languages which are classified as Amerind. Critics often point out errors in transcription and other minor details to suggest that Amerind is a poor quality comparison although few have the evidence which would suggest the whole work was not credible. In fact Greenberg
summarizes his article *In Defense of Amerind* by quoting Lamb who writes, "The volume calls for careful study and for follow-up research. It does not call for criticism based on incomplete understanding of his methods" (1987).

Another proto-language which is on the same level as Amerind is the proto-language of Nostratic which is a term that is somewhat ethnocentric meaning "belonging to us/to our part of the world i.e. Eurasia plus Northern Africa" (Hock and Joseph 1996). The validity of Nostratic, like Amerind, is also not accepted by the majority of historical linguistics, but "if it is not proven, it is not discredited either" (Nettle 1999). Again, it is the lack of a concrete methodology which gives issue to the controversy over Nostratic, and similarly to Amerind, Nostratic's biggest critics are convinced that the Proto-language contains too many languages, and that the similarities between them are just random. However, unlike Amerind the languages of Nostratic are spread across many different regions, and because the origins of people in these regions are not the result of a migration in fixed time and place like the arrival of people into the New World critics of Nostratic claim that the combination of the language families is less accurate. The original compilation of Nostratic was also completed by one scholar, Aharon Dolgopolsky, and this is also cause for concern as many other linguists including Lyle Campbell find serious problems with "nearly all of Dogopolsky's 124 Nostratic lexical sets" (1999). There are still many doubts about the soundness of Nostratic, and unless it can be supported with more evidence linguists will continue to have these doubts, and "any interpretations of non-linguistic prehistory based on it remain sheer speculation" (Campbell 1999).

Use of non-linguistic sources to corroborate these linguistic comparisons may be the way to eliminate these doubts from the minds of many scholars especially when these other sources might provide concrete chronologies and observable changes throughout evolution of people which historical linguistics can not see. Raimo Antilla writes that "collaboration between archaeologists and linguists is desirable because each side tends to apply the findings of the other too simplistically" (1989).

**GENETIC EVIDENCE**

For a long period of time there has been the hope that "evidence from molecular genetics will cast more light upon population histories which may in turn have a bearing on language history also" (Renfrew 1998). The use of genetic evidence for linguistic comparisons is discussed in this section. Genes and languages have many similar traits in common when it comes to the way in which they spread, and the ways in which they are studied consequently. Both utilize tree diagrams in an attempt to relate the genes/languages of today back to an original source as well as show the "representation of inherited relationships" (McMahon, Lohr, and McMahon 1999). One fundamental difference which McMahon, Lohr, and McMahon note in their article "Family Trees and Favorite Daughters" is that biologists use trees that are more based on quantitative data while linguists work more intuitively (1999).

Cavalli-Sforza has been the main biologist who has done studies which link these two separate phenomenon together. He suggests that the spread of genes and languages are linked together, and that studies of mtDNA enable patterns of gene diffusion to be isolated to individual populations. Thus, if a gene is in one population, and moves to another, but not to a third then those populations that
have the same gene must also have similar languages, and consequently cultures. Cavalli-Sforza also notes that there are fewer linguistic trees than genetic ones, but "it is usually true that the genetic similarity between populations belonging to the same linguistic family is high" (1997). There are a few exceptions to this rule including the populations of Lapps, Ethiopians, and Tibetans each of which speak a language to which they are not biologically linked. These exceptions may be related to gene flow and language replacement, and while they "blur the genetic and linguistic picture...they do not obscure it entirely" (Cavalli-Sforza 1990).

Cavalli-Sforza also considers the place of origin of the first populations of anatomical moderns by examining how closely related populations are to one another. Cavalli-Sforza indicates his belief that further collaboration between genetics and linguistics will be beneficial in examining the origins of modern humans. There are similar problems within each discipline as the rate at which genes and languages change is still controversial, and the rate at which genes change is one of the main differences between the continuity and replacement models of human origins.

In respect to this genetic evidence Cavalli-Sforza indicates that all populations are related, and that Africa and Asia are more closely linked than Europe (1997). As well, one of the earliest branches in the evolution of modern humans is the separation of Africans and Non-Africans (1997). This suggests an African Origins Model, and would therefore suggest a replacement model. Linguistic evidence also suggests a replacement model (Ruhlen 1994b) as it is with the arrival of anatomical moderns that the cognitive capacity for language is present suggesting monogenesis of all languages. Genetic studies of specific areas which link linguistic groups to populations are also used in attempting to prove comparative linguistic studies. These include theories by Turner on dental evidence, and Zegura on genetic differences such as blood types which indicate that there were three distinct migrations into the Americas, and the populations of these migrations correlate with Greenberg's proposal of three linguistic groups of Eskimo-Aleut, Na-Dene, and Amerind (Greenberg, Turner, and Zegura 1986). Similar work has been done which connects Nostratic to genetic populations of Eurasia and North Africa; the areas from which the languages of Nostratic are taken (Barbujani and Pilastro 1993). There have also been demonstrated genetic and linguistic population convergences in Australia (Ruhlen 1994b), and Southeast Asia (Baer 1995).

**FOR AND AGAINST PROTO-HUMAN**

The current view of most historical linguists is that there is a time-limit past which languages can not be reconstructed. This limit has been suggested at different times which vary drastically from 5,000 (Hock and Joseph 1996), to 6,000 (Ruhlen 1994a; Ehret 1999), to 7,000 years (Greenberg 1996) depending on the study. All this variation between different scholars beliefs of when linguistic comparisons can no longer be done appears to indicate uncertainty in the field, and suggests that perhaps the time-limit is not so early as it was once thought to be. Few linguists would admit to this, however, who do not already believe in the deep-time comparisons of languages leading to proto-languages such as Nostratic and Amerind. Some linguists such as Ruhlen and Greenberg, as well as a few others, suggest taking comparisons...
of language even further than this to create what might be known as proto-human where the "ultimate goal [is] a comprehensive classification of what is very likely a single language family" (Greenberg 1987). The construction of this proto-language is situated in the belief that:

Human language came into being just once and that all languages that now exist (or ever have existed) are (or were) altered later forms of this original language" (Ruhlen 1994b).

Greenberg and Ruhlen have even suggested 27 global etymologies which may in fact be evidence of proto-human, and Ruhlen suggests that these are only a fraction of what could actually be found if closer comparisons were made (Ruhlen 1994a).

It is my belief, however, that eventually a point is reached where proto-languages come to the extent of the data that the languages of present provide, and any other comparisons beyond this are too far away from the original data to be correct. To summarize, the words of the proto-language are so distorted that it is impossible to tell whether the connection between cognates is a true one or not because we will never know if the proto-languages are even close to languages which were spoken in the past. This belief is supported with data in Hock and Joseph's (1996) book which suggest that the global etymologies of Proto-human as suggested by Greenberg and Ruhlen are "overly short". This is problematic in the sense that chance similarities are more likely to occur in short words than in longer ones (Hock and Joseph 1996; Moravcisk 1978). The one etymology which is not too short is the proto-human term maliq'a meaning throat or swallow (Hock and Joseph 1996). Greenberg and Ruhlen find this cognate in "Amerind, Eskimo-Aleut, and four language families of the Old World: Afro-Asiatic, Indo-European, Uralic, and Dravidian" (Hock and Joseph 1996); it appears in thirty-two languages in all. Even with this suggested similarity in many languages numerous linguists find problems with the semantics of the cognate as they include such varying meanings as, "swallow, throat, suck, chew, milk, breast, and neck" (Hock and Joseph 1996).

Other indications that these global etymologies might be incorrect is seen in the article written by Trask which states that he can find fifteen similarities between Basque, a reputed language isolate, out of the 124 cognates described in the proto-language of Nostratic (1999). If a language which is known for being isolated linguistically, and genetically (due to the highest frequency of Rh-negative gene of any population (Ruhlen 1994b)) can be seen to have similarities with higher level proto-languages the legitimacy of this type of study for deep time-frames is called into question. It is also important to note that the Basque language survived the migration of Indo-European languages into its region (Ruhlen 1994b). This suggests that Basque is the exception to the rule that languages spread with people, and this challenges the assumption of language replacement in all territories except previously non-populated ones (Renfrew 1990).

Pessimistic views of this sort which doubt whether there is enough evidence to support proto-human languages have been around for a long time. As early as 1867 American historical linguist William Dwight Whitney wrote, "Linguistic science is not now, and cannot ever hope to be, in a condition to give an authoritative opinion respecting the unity or variety of our species". This is not to
say that linguistics, archaeology, and genetics are not beneficial to each other but the evidence which linguistics can provide is set within a limited time-frame. As Ruhlen writes "we know a good deal about what our ancestors looked like, and how they lived, but their minds and languages remain shrouded in the past" (1994b).

**CONCLUSION**

This paper has discussed what comparative linguistics are, and the timeframe in which they can be applied, as well as why it is highly unlikely that use of linguistic evidence can be traced back as far as the origins of modern humans. This essay argues that linguistic evidence is useful, however, in showing that populations of today are related, and how these changes in languages have occurred as populations have spread across the globe. The languages of Amerind are especially useful to study as they show how languages spread across a previously uninhabited continent (according to archaeological evidence), and at what rate. Although the true diversity of all languages spoken can not really ever be known approximate rate of change is an important key to determining how far back languages might be related even if they cannot physically be connected. This essay has also looked at the place of genetics in collaboration with linguistics, and how the combination of the two pieces of data provide for a fuller understanding of how anatomical moderns spread at later time-periods than 15,000 years ago. Whether or not Ruhlen's notion of a proto-human language is a reality or not is difficult if not impossible to determine using the evidence of the present. I quote Colin Renfrew's current appraisal of the evidence to date on the use of linguistics to prove time and place of human origins:

> If the lumpers (notably the Russian school including Dolgopolsky, and the American school of Greenberg and Ruhlen) are correct, then the unity of human origins may be reflected in the evidence offered for linguistic monogenesis...But if the splitters are right, and if it is indeed the case that languages evolve so fast that no reconstructions of any kind could be possible beyond about 5,000 years ago..., then, these broad macro-families (e.g. Amerind, Austric, Nostratic or Eurasian, Sino-Caucasian) would be entirely illusory, and so too in consequence would any supposed correlation between them and the genetic evidence for human phylogeny... (1992).

I personally believe that a single origin of all human languages does exist, but trying to prove its existence is another matter as is attempting to describe it. The data eventually gets too thin to postulate a reliable version of a proto-language which encompasses all of the world's languages and consequently the language of the original anatomically modern humans.

**REFERENCES CITED**


Campbell, Lyle 1999. "Nostratic and linguistic palaeontology in methodological Perspectives." In *Nostratic: Examining a


Schreyer – Proto-Human Language 60


--- 1990. “Archaeology and Linguistics: Some Preliminary Issues.” In When Worlds Collide: Indo-Europeans and Pre-Indo-


