

Going to the Root: Paving the Way to Reconstruct the Language of Homo-Sapiens

Evangelos C. Papakitsos¹ & Ioannis K. Kenanidis²

¹ Department of Education, School of Pedagogical and Technological Education, Greece

² Primary Education Directorate of Kavala, Greece

Correspondence: Ioannis K. Kenanidis, Primary Education Directorate of Kavala, Ethnikis Antistaseos 20, 651 10 Kavala, Greece. E-mail: ioakenanid@sch.gr

Received: April 13, 2018; Accepted: May 4, 2018; Published: May 9, 2018

Abstract

This paper presents an attempt to reconstruct the most basic features of the language of Homo Sapiens, following the principle of monogenesis, namely the viewpoint that since humans share a common biological ancestry, they also share a common linguistic one. Considering this issue, the basic methods of comparative linguistics are briefly presented first, along with the methodological approach utilized herein, named Qualitative Inquiry. The results of the reconstructing process are presented, classified in terms of phonological, morphological, lexical, grammatical and syntactic aspects. Only bordering to the scope of this paper, a brief comparison of this treatise to previous studies reveals both convergence and discrepancy concerning the features of the language.

Keywords: monogenesis, language reconstruction, Proto-Sapiens, comparative linguistics, Qualitative Inquiry

1. Introduction

When L. L. Zamenhof constructed Esperanto (Kiselman, 2008), he thought that people cannot understand each other because they do not speak one common language, putting the cart before the horse: in reality, people do not speak one common language because they do not want to understand each other. Everywhere we may look at, we find people grouping together in order to fight against other groups of people. We reason that it was not so with our ancestors: the early humans had to unite in order to survive amidst formidable beasts, temperatures much below freezing point, scarcity of food and deadly diseases. This is only one of many considerations reinforcing the theory of monogenesis (Kozyrski & Malovichko, 2015), which is the deduction that all human languages originate in one proto-human language: A theory that has given rise to much controversy, being opposed by the theory of polygenesis (Jackson & Weidman, 2005).

There are linguists who reject altogether the idea of reconstructing or even describing a single proto-human language, as being “a hopeless waste of time” and “an embarrassment to linguistics as a discipline” (Campbell & Poser, 2008, p. 393), arguing that so much time elapsed since the hypothesis that such a language existed. On the other hand, the common ancestry of Homo Sapiens is an anthropological theory supported by strong evidence, especially regarding the last migration wave out of Africa, estimated approximately from 70 to 50 thousand years ago (López et al., 2015; Endicott et al., 2007; Metspalu et al., 2006). This last group of immigrants, whose descendants eventually inhabited our entire planet, is estimated to have been no more than 1,000 individuals, originating from a larger group of no more than 5,000 individuals (Stix, 2008; Zhivotovsky et al., 2003). On these grounds, it is reasonable to assume that the community of all those people was a monolingual one, thus the attempts to reconstruct or describe their language is no less sensible than trying to reconstruct or describe any other proto-language (Rowe & Levine, 2015, pp. 340–341; Koerner, 1999, p. 109), as for example the Proto-Indo-European (Lehmann, 1993).

In the frame of the aforementioned debate, the present study focuses on the language of the mentioned group of humans who left Africa, from 70 to 50 thousand years ago, to inhabit the rest of the planet, carrying with them a symbolic culture and a language (Stringer, 2011). Issues prior to that era, regarding the very origin of speech (Hockett, 1960), language, cognition and their related theories (Ulbæk, 1998) are not the focus of the research presented here, since such issues (not directly concerning the scope of this work anyway) are considered to be among the hardest problems of contemporary science (Christiansen & Kirby, 2003). For that common ancestral language, several names have been used occasionally: Proto-World (Fleming, 2003), Proto-Human (Bengtson, 2007) and Proto-Sapiens (Ruhlen, 1994a). The last term will be preferred here (Proto-Sapiens), because the first

one could refer to hypothetical languages even before the emergence of humans (Hamrick et al., 2018), while the second one ignores the potential existence of languages of other human species, like the Neanderthals (Lieberman, 2007; Arensburg et al., 1989).

The reconstruction of a proto-language is conducted through the comparative method within a group of languages with similar features, eventually leading to the formation of a linguistic family tree (or “phylogeny”). For coping with various incompatibilities of phylogeny, the wave model can be additionally or alternatively applied (Alexandre, 2014; Heggarty et al., 2010; Labov, 2007). The comparative method has been accused of being “an intuitive undertaking” (Schwink, 1994). However, besides having many words (currently 1152 “long range” and 2077 “Nostratic” lemmata) reconstructed from linguistic families’ proto-languages recorded in The Languages of the World Etymological Database (LWED) that is part of the Tower of Babel project (Starostin & Bronnikov, 1998–2009), which were reconstructed from actual words and not from intuition, a more recent study indicates a common linguistic ancestry of proto-languages across Eurasia, by using modern quantitative modeling (Pagel et al., 2013). According to this study:

- words that are used in everyday speech more than once per thousand are 7-10 times more likely to show a common linguistic ancestry in the Eurasian proto-languages;
- numerals, pronouns and special adverbs exhibit a linguistic half-life of once every 10,000 years or even more, thus being replaced far more slowly;
- 23 words are identified to have survived 15,000 years (Note 1);
- the previous words indicate the existence of a linguistic common ancestry for seven proto-languages of Eurasia (Note 2).

Accordingly, there must have been a proto-language spoken in Eurasia 15,000 years ago.

Considering the specific results of reconstructing words of proto-languages, the relevant critique is focused on the applied methodology concerning the accuracy of the reconstructing process. A brief description of the reconstructing methods will be presented next, along with the approach that is followed herein.

2. Methods

The methods, that will be briefly commented on for languages’ reconstruction include: the Comparative Method; Mass Comparison; Lexicostatistics; Internal Reconstruction and Glottochronology. Finally, the methodology that we name Qualitative Inquiry will be introduced in this work.

2.1 The Comparative Method

The Comparative Method (alias “traditional” or “standard”) in comparative/historical linguistics is considered the predominant method for determining the common linguistic ancestry of two or more languages, by comparing their features. In modern times, it was originally developed during the 19th century, by scholars who had studied the common descent of the Indo-European (IE) linguistic family (Lehmann, 1993; Schleicher, 1874) from the Proto-Indo-European (PIE) language. The comparative method includes the comparison of lists of cognate terms along with sound correspondences. Then, the corresponding proto-languages may be reconstructed through the study of regular changes of phonemes through time and regular phoneme correspondences across cognate languages, in accordance, of course, to “sandhi” or “conditioning environments” such as word stress, accent or other adjacent phonemes. The suggested set of steps in applying the comparative method (Campbell, 2004; Crowley, 1992), although not unanimously accepted by the historical linguistics scientific community, are presented here:

- The assembly of potential cognate word lists, containing the basic words’ meanings of the examined languages (Lyovin, 1997).
- Establishing the corresponding sets of words.
- Discovering those sets that are in complementary distribution (Note 3).
- The reconstruction of the equivalent phonemes of the proto-language.
- The typological examination of the reconstructed phonetic system.

After determining a group of languages originating from one reconstructed proto-language, a family tree is designed (“tree-model”), depicting the linguistic ancestry through language evolution (Lyovin, 1997).

The main limitations of the Comparative Method can be summarized in the following points:

- The initial theoretical hypothesis that the phonetic laws, causing sound changes from one language to

another, have no exceptions is criticized due to the fact that many words are observed to change in irregular manners. An alternative revising hypothesis is that “each word has its own history” (Szemerényi, 1996).

- Any study should carefully take into account the word-borrowing from other languages. Consequently, for the purposes of comparison, it is mainly the basic vocabulary that should be used, such as pronouns, body parts or small numbers (Dolgopolsky, 1964), that is words resistant to borrowing (Bengtson & Ruhlen, 1994).
- Any reconstruction of a linguistic family-tree should deal with potential “area diffusion phenomena”, such as a “Sprachbund”, that may cause a false classification (Dixon & Aikhenvald, 2001).
- The consideration of convergence, concerning motivated or accidental resemblances between words of different languages, is very important (Bengtson & Ruhlen, 1994).
- The tree-model of a language family should not be regarded as a uniform or well defined depiction of diachronic language evolution, for several reasons, especially due to dialect divergence (Campbell, 2004).

Dealing with the tree-model criticism, the “wave-model” has been alternatively proposed for depicting language evolution (Fox, 1995). The wave-model uses Venn diagrams for the representation of languages’ change that may intersect, unlike a tree-model.

2.2 Mass Comparison

The method of Mass Comparison is considered incompatible with the Comparative Method by most historical linguists (Campbell, 2001), although its advocates reasonably hold a different opinion (Bengtson & Ruhlen, 1994). There are several other terms used for this method: “mass lexical comparison” (Greenberg, 1957), “multilateral comparison”, “taxonomy” or “classification” (Ruhlen, 2005).

Mass Comparison had been used by Greenberg (1957) for determining language families. This method involves the initial construction of lists of basic words (e.g., body parts, small numbers, pronouns or common morphemes) from many languages geographically related (e.g., African). Then resemblances are sought for sound correspondences to discover languages that could be related. Once selected, the Comparison Method (see subsection 2.1) is applied to these languages for discovering a potential common ancestry. In this respect, the detractors of Mass Comparison (e.g., Campbell, 2001; Hock & Joseph, 1996) claim that this method doesn’t conduct a systematic study of lexica, morphologies and grammars of the examined languages. On the contrary, its defenders (Greenberg, 2005; Ruhlen, 2005) claim that this is merely the first stage before the application of the traditional method (see subsection 2.1). Ruhlen (2005) accurately observes that the first two stages of the Comparative Method, as stated by Durie & Ross (1996), are given in reverse order. These are:

- determining a set of languages to be examined as potentially related, through “diagnostic evidence” and
- collecting putative cognate sets for that linguistic family,

The first step is based on “mysterious diagnostic evidence” for identifying a language family, while the second step actually looks for the required lexical or grammatical evidence. Thus, the order should be reversed so as to apply Mass Comparison that will collect cognates’ sets for a group of languages and then determine which of them could be related, before continuing with the application of the rest of the Comparative Method.

2.3 Lexicostatistics

Lexicostatistics is a method developed by Morris Swadesh (Ruhlen, 1994b; 1994c; Hymes, 1960) for determining the relationship between languages by comparing the percentage of their lexical cognates through a mathematical equation. This method follows the next steps:

- A list of universally used meanings is created (body parts, pronouns, natural themes), having about 200 of them (Swadesh, 1955).
- For each language of an examined group, the words corresponding to those meanings are collected.
- The correspondent terms across languages are determined and tagged as positive, negative or indeterminate, from the aspect of their apparent cognateness.
- The lexicostatistic percentages are calculated to the proportion of words that are cognate for each pair of languages.
- Every value of lexicostatistic percentage is entered into an $N \times N$ table of distances, where N is the number of the examined languages.

- Based on the proportion of cognateness from the previous table, a language family-tree is created.

The criticism to this method mainly refers to the feasibility of creating a list of universally used meanings (Gudschinsky, 1956).

2.4 Internal Reconstruction

The method of Internal Reconstruction examines the evolution of a single language from past forms to more recent ones (Givón, 2000). It can be used as an auxiliary means to the previous methods, especially when trying to determine the proto-language of a family tree, potentially spoken many centuries ago.

2.5 Glottochronology

The method of Glottochronology is regarded as a part of Lexicostatistics (see subsection 2.3) that deals with the time elapsed in word changing (McMahon & McMahon, 2005; Embleton, 1992). It has been also developed by Morris Swadesh (1972, pp. 271–284), using mathematical formulas analogous to radioactive decay. There are various issues that this method has been criticized for, but also many cases that have demonstrated its validity (Renfrew et al., 2000), while similar methods or modifications have been also developed recently (Gray & Atkinson, 2003; Starostin, 2002).

2.6 Qualitative Inquiry

The methodology of Qualitative Inquiry (Kenanidis, 2013) has been more used for reconstructing forms and rules of the Proto-Sapiens language rather than of newer languages. It comprises the following specific principles and guidelines, within the overall framework of comparative linguistics:

- If there is a common anthropological ancestry of Homo Sapiens, then there is a common linguistic ancestry, as well. In this respect, there is no such thing as “language isolate”. Even languages generally considered “isolates” can offer some help in reconstructing the ancestral language, although, of course, older and conservative languages offer more and better material for that purpose. The criterion for language conservativeness is etymological transparency: the more easily we can find the roots that the words belong to, the more conservative is the language’s phonology.
- “Words are many but roots are few”: Comparative linguistics is a serious science, when we compare roots and not words. We seek the root within each language before we compare the roots of different languages. Words, even of not “basic” meanings, are valued for comparison to the extent they help in determining a root; otherwise they should not concern comparative linguistics. Comparing words can be greatly misleading, but when we go to the root, comparison is enlightening (see Appendix A).
- The relationship between signifier (word root) and signified (concept) was neither conventional nor arbitrary (e.g., see: Yi-chün Lu, 1998); there was a reason why this root form had that meaning and not another one, and our common ancestors knew that reason. Therefore, each phoneme and consequently each phoneme combination had an unchanging meaning known and felt the same by all people. It has only sporadically been noticed that some phonemes in ancient languages seem to have a meaning on their own, and still today there is only a twilight of a serious research that may explain how the individual phonemes stood as the core of the lexical roots (for examples see Appendix B).
- For the reasons of the previous paragraph, there are roots cognate to each other too. An interesting example is PIE root “tjeg” (= to hold back, withdraw), reconstructed from Greek seb-, Sanskrit “tyag”, etc., and the Proto-Turkic root “tjak” (= to go out or climb up, as when avoiding a danger) reconstructed from Turkic çık- /tʃuk/. The vowels did not differentiate roots, but the Turkic root must have been with /k/ and not /g/, since an Old Turkic /g/ is regularly preserved as ğ/ɣ/. Another example are PIE *mel (Greek “meli”, Latin “mel”: honey) and *medhu (Greek “methu”, Sanskrit “madhu”: originally a kind of sweet wine made from honey).
- Comparing supposedly unrelated languages also reveals grammatical morphemes of the ancestral language, which originally were lexical morphemes. For example, Sumerian, Turkic, Austronesian and Semitic languages are considered unrelated and very different to each other, however all those languages show an affix “ma” (suffix in Turkic, prefix in Austronesian and Semitic; Sumerian has turned it to “ba-” following a usual phonetic tendency), which means the thing resulting from an action. So this affix “ma” is also related to the PIE and Proto-Uralic accusative suffix –m (because the accusative indicates the result of the verb). And of course, -ma, inflected as -men in Latin, -ma(n) in Sanskrit, -ma(t) in Greek, is a very well-known PIE suffix for things resulting from actions.
- The Qualitative Inquiry asks certain questions, even if that may seem tedious: Is this word a form of one

and only root? Or does it contain inflectional or derivative affixes? What remains if we remove the affixes? What phonetic alterations could it have suffered, under what phonetic laws, in this particular language? Are there any older forms of this word (or its root) in older stages of this language? What other words are derived from the same root in this language? Are there any apparently related (cognate or borrowed) words in other languages of known affinity of possible proximity? It is only by asking these questions that we can proceed to detect a root and then compare it to a root in a different linguistic family.

- To determine whether a word is a loanword or not, we need to know how the different languages could have come into contact, so that the word could be borrowed. If it was impossible for the languages to come into contact, then we cannot think of word-borrowing, so we must think of cognateness. Of course, the type of word can indicate whether it could be a loanword or not: a thing that came from another culture is likely to be named by a borrowed word, as the Turkic “bö:z” came indeed from the Greek “buss-os” (= a type of precious cloth). But a word of a very native meaning, like the Arabic “wadi” or the Australian Western Desert (WD) aboriginal “karu” (both meaning a temporary rivulet), cannot be borrowed from a distant language. Phonology is very helpful in revealing loanwords because every language “simplifies” to its own means what it borrows from another language: when we have PIE **“suh”* and Proto-Turkic */*εox/* (both meaning “domestic pig”), we understand that PIE borrowed the word, because Proto-Turkic had /s/, but PIE did not have /e/ or /j/. Morphology is also useful for the same purpose: Turkish “günder” shows no Turkic means of derivation, but the original Greek “kontari” (/kundär/ by northern Greek phonotactics) is formed with the very common Greek suffixes “-ar-i”.
- Cognate words in different linguistic families are generally not expected to look very much alike and often they are hard to be noticed, unless the phonetic laws that have altered them in each language are known, e.g., Turkic “aşuk-” (= hope) and Greek “elp-” (= hope) appear quite different, but knowing the phonetic evolution in each language, a Proto-Sapiens form */a|q/ is readily revealed. So, whenever words resemble each other too much by both form and meaning, in languages separated by many centuries and many phonetic changes, then the resemblance is suspected to be fortuitous (for examples see Appendix A). On the other hand, a similarity is only fortuitous if it is proven to be fortuitous by means of showing the different origin of similar words. For example, the Modern Greek “mati” comes from “ommatation”, where only “om-” was the root and even that “om-” came from “op-” and from PIE */*seq/-*. So, it is proven that Greek “mati” has nothing to do with Austronesian “mata”. However, if there is no such proof, then the similarity reveals a real cognateness due to the common origin of languages. Many academicians, as Clouston (1972) did, reject a cognateness without giving any explanation of similarities, like *kör-* and *kara-* (see next paragraph), and even between words that have to be original, like basic pronouns. This is prejudice, a factor that has greatly impeded the progress of comparative linguistics.
- Some phonemes are very easy to change with time, for example a /*η*/ especially in the beginning of words, while other phonemes are very durable and can be found similar in the most diverse languages, as for example /*c*/ in Latin “cerno” (= to see), Greek “krinō” (= to discern, estimate), WD aboriginal “kuru” (= eye), Turkic *kör-* (= to see), Mongolian *kara-* (= to watch). According to Clouston (1972), Mongolian is not genetically related to Turkic, only there are many loanwords mainly from Turkic into Mongolian. Indeed, as Clouston (1972) observed, Mongolian *kara-* was borrowed into Turkic (and then into Modern Greek in the word *καραούλι* /*kärä.uli*/), but that *kara-* in Mongolian is not a loanword: it is simply cognate to all the other words mentioned, in languages supposedly unrelated. So, it often happens that a Proto-Sapiens root has given words both directly in a language and indirectly through borrowings from another language (see Appendix C).
- National pride is good, as long as it does not get to be blinding nationalism. Although no researcher wants to be blinded, such instances sometimes have occurred. All etymological dictionaries of Turkic languages and of IE state that Turkic “bal” (= honey) is an early loanword into Turkic from some IE language; the same is stated for Chinese 蜜 “mjēt” (Kalgren, 1923: entry 617). But how do we know that these IE words (“mel” and “medhu”) were not borrowed in the reverse direction, which is from Turkic or some other oriental language into PIE? We have evidence that PIE borrowed some important cultural words from ancient Turkic (Appendix C), relating to words for domesticated animals and metals. If the very name “Turkic” is associated with Indo-European contempt, then it was not Turkic, but an ancient “Altaic” language that lent those words to the PIE language, along with important cultural innovations.
- Words carry history and culture. Sumerians called themselves “the black head (i.e., black haired) people”. The term for “head-black” is reconstructed as **twop-nejec* and that is considered cognate to the Finnish term “Suomi”, possibly even to the Aztec ethnic group “Tepanec”, and equivalent to the Turkic autonym

reconstructed as **twep-quro* (Kenanidis, 2013), hence “Törkü”, nowadays “Türk”. But how did they notice that they were black haired if they didn’t know of neighboring people with blond hair? And why would they designate themselves as “black haired”, if that did not distinguish them from neighboring people with blond hair? Although many loanwords have been found from ancient Turkic into PIE, and a few questionable loanwords in the opposite direction, that self-designation of Sumerians and Turkic nations alone is enough to prove that they were neighbors to a fair haired nation, and that themselves were uniformly black haired in ancient times.

The application of these principles and guidelines resulted in an approximate reconstruction of the Proto-Sapiens language features, as presented in the next section.

3. Results

A few Proto-Sapiens words will be presented as discovered by the method of “Qualitative Inquiry”. The Aztec language is natively called “nahuatl”, meaning “speech, language” (literally: “clearly understood sounds”). The same word (“nahuatl”) is also found in some aboriginal languages of South Australia as “*ḡawa|a*” (= speech, language). The Australian aboriginal languages, being separated from the rest more than 50,000 years ago, retain words and sounds that were changed in most languages, such as the original *ḡ-* that the Aztecs turned to *n-*, the *w-* that became /*m*/ in Nahuatl (written as *-hu-*) and the retroflex /*l*/, corresponding to the very common Nahuatl “*tl*”. So, the reconstructed word /*ḡäwä|*/ (= language) was an autonym of Homo-Sapiens language (Kenanidis, 2013). An approximate proposed reconstruction of Proto-Sapiens (“/ḡäwä|”) will be presented in terms of Phonology, Morphology and Lexicon, Grammar and Syntax (Kenanidis, 2013; 1992).

3.1 Phonology

Proto-Sapiens was exceptionally rich in consonants, while it was poor in vowels, which, anyway, stood for inflexion only and could not differentiate roots. A special category of consonants were the liquids (rhotics and laterals), which were used extensively, enriching words with meanings and feelings more intensely than the other consonants. Every obstruent consonant was necessarily followed by a vowel, but the continuants (including the liquids and nasals) were excluded from this rule.

Apart from the liquids, the articulatory positions of the other consonants were four (4): soft (back) palate, hard (front) palate, teeth (including alveoli) and lips. There is a question whether retroflex consonants formed a distinct articulatory position. Our answer is “no”, given that the retroflex consonants, apart from the liquids and the relative sibilants, appear almost exclusively in Australian and Dravidian languages, where articulatory manners were oversimplified. So, the retroflex consonants there seem to exist only to enrich the phonemic inventory, and even so they only exist after vowels. Thus, we hold that the articulatory positions were only the four mentioned, with a possibility to pronounce a retroflex being only an allophone of the emphatic dental, but retroflex liquids (lateral and rhotic) did exist as distinct phonemes in Proto-Sapiens.

In each of these four positions (velar, palatal, dental and labial) there were a plain stop consonant, an emphatic stop, an aspirate stop, an emphatic aspirate stop, a nasal, a fricative, and an approximant. The “emphatic” were mostly realized (probably) as geminated (e.g., “pp”) or tense, like the Korean “ㅃ”, but the same could also be realized as pre-nasalized consonants, e.g., “mp”. While the simple aspirate consonants are well documented by PIE, there is hitherto little only evidence for the emphatic aspirate consonants: they are documented, for example, by Greek “*enkhos*” (= spear), corresponding to the Turkic “*süḡü*”, probably from Proto-Sapiens **sjeqqho* / **sjeḡqho*. A word for blood was probably **saqqh-*, hence the Latin “*sanguis*” and Greek “*hai-ma*” from **saqqh-ma*; North Tungusic Oroqen: /*ʃə:kʃə*/, South Tungusic Manchu: “*senggi*”.

The approximant velar, palatal and labial were approximately “*u*”, “*j*”, “*w*” respectively, while the approximant of the dental position must have been a kind of “*s*”. The dental fricative was /*θ*/, therefore quite different to “*s*”. The “*s*” itself had an allophone of “*r*”, even before the Proto-Sapiens language split into different languages: some people and groups used “*s*” while others used “*r*” in its place. The correspondence between “*r*” and “*z*” has been preserved especially in Turkic languages (which are classified as *z*-Turkic and *r*-Turkic), but it can be found in most linguistic families too. From the correspondence of sibilants to liquid consonants, as found in the languages known today, we judge that the western parts of ancient nations preferred the “*r*” while the eastern parts preferred the “*s*”. Some Amerindian languages lack liquid sounds altogether, while they are rich in sibilants: those were eastern Asian tribes that travelled eastwards until they crossed the Bering Strait and settled in America. The Australian aboriginal languages (rarely having some fricatives) do not have “*s*” at all, while they keep the richness of liquids that existed in the Proto-Sapiens language. This is another indication that Proto-Sapiens did not have sibilants except a “*s*” being an allophone of “*r*”. Also, there was no voicing distinction: apart from semivowels,

liquids and nasals, all consonants were normally voiceless and any untypical voicing could not change their meaning.

Because the exact nature of each Proto-Sapiens phoneme cannot be known with accuracy and, anyway, each phoneme could have dialectic and allophonic variants, also in order to avoid strange symbols and diacritics as much as possible and to be in accordance with ways of transcribing words of various languages in other publications, we do not use exactly the International Phonetic Alphabet (IPA) for the Proto-Sapiens language, but a specific, non-case-sensitive usage of the alphabet as shown in Table 1 and Table 2.

Table 1. The (non-liquid) consonants of Proto-Sapiens

<i>Approximate position</i>	<i>Plain</i>	<i>Emphatic (with secondary forms)</i>	<i>Aspirate</i>	<i>Emphatic Aspirate (with secondary forms)</i>	<i>Nasal</i>	<i>Fricative (estimated as:)</i>	<i>Semivowel (approximant)</i>
<i>Velar</i>	Q	QQ (ŋq)	QH	QQH (ŋqh)	Ḍ	H (/x/)	ɥ
<i>Palatal</i>	C	CC (ñc)	CH	CCH (ñch)	Ḍ̃	Š̃ (/ç/)	J
<i>Dental</i>	T	TT (nt)	TH	TTH (nth)	N	Θ (/θ/)	S/R
<i>Bilabial</i>	P	PP (mp)	PH	PPH (mph)	M	F (/ɸ/)	W

Table 2. The liquid consonants of Proto-Sapiens

	<i>Palatal</i>	<i>Retroflex</i>	<i>Alveolar trill</i>	<i>Alveolar/Dental</i>
<i>Rhotic</i>		,R	RR	R (approximant, /ɹ/) alternating with S
<i>Lateral</i>	Ł	,L		L

Because the bi-consonantal roots very often included a semivowel (or liquid) between the two main consonants, all consonants could be followed by J, W, ɥ, and possibly also S/R. This resulted in an even greater variety of consonants, which was mostly lost in the descendant languages. However we know that such a rich variety did exist, as extraordinarily numerous seemingly identical roots in various linguistic families have unrelated meanings (derived from different Proto-Sapiens roots, originally differentiated through different consonants). Especially, we know with certainty that the Proto-Sapiens language had a rich inventory of rhotic and lateral phonemes, because similar inventories are found in all Australian aboriginal languages and also because extraordinarily many seemingly identical roots with unrelated meanings are found in all natural languages, which roots contain rhotic or lateral sounds and no difference can be found in the rest of their consonants. There is no available evidence for any uvular or laryngeal rhotic/lateral consonant, which are actually difficult for most people to pronounce. Some reconstructions of Proto-Austronesian (Zorc, 1995; Blust, 1999) include a /r/ (uvular R) but we find that this is a misconception (see Appendix D).

Proto-Sapiens distinguished only three vowels, “E” (front), “A” (central) and “O” (back and rounded). The vowels were quite opener than /i/, /ə/, /u/ respectively, so as to avoid confusion with the semivowels /j/, /ɥ/, /w/. Still an /ə/ might have been used as an epenthetic, without any meaning, after words ending in obstruent consonants, only to enable such consonants to be pronounced. The ordinary vowels {E, A, O} could be lengthened so as to express the speaker’s feelings: a longer vowel indicated a stronger feeling, which meant emphasis on the size, duration, distance, strength (etc.) of the thing spoken of. For convenience, we represent long vowels by reduplication.

3.2 Morphology and Lexicon

The roots of the Proto-Sapiens language were of the C(S)C type, where C = consonant, S = semivowel or liquid consonant; also, C (single consonants) were meaningful roots. Single vowels were only used as demonstrative (and consequently personal) pronouns.

Tri-consonantal roots are known (and common) in the Semitic languages, but note that the other languages of the Afro-Asiatic family (which the Semitic belongs to) use mainly bi-consonantal roots, which confirms that the Proto-Sapiens roots were essentially bi-consonantal (consisting of two main consonants); but the bi-consonantal roots

were in most cases enriched by a semivowel or a liquid sound. When the semivowel or liquid was lost, the Semitic languages replaced it with a third consonant, while the other Afro-Asiatic languages did not replace it (Note 4). For example, the Proto-Semitic root g-r-b (= foreigner, strange) must have been expanded from a Proto-Sapiens root *qq-w-r, attested in Sumerian “kúr” (= stranger, not one’s own) and Latin “curius” (for other examples of Semitic roots formation see: Appendix E).

It appears that meaning was primarily connected to articulatory positions, and even vowels were defined by articulatory positions (velum for A, palate for E, lips for O); but vowels, being only air blowing, were only used to express the speaker’s feelings and attitudes, and therefore grammatical concepts, not lexical meanings. The other means to express what we call today “grammatical concepts” were the word order and intonation. All these that later became “grammatical morphemes” were originally ordinary words.

This structure of the Proto-Sapiens language, with any slight change of sound making a difference in meaning, made it a language extremely resistant to change. A comparable obduracy is observed in Semitic languages, because there too vowels did not differentiate roots and any change to the consonants would cause a change of meaning, thus making communication impossible. So, the Proto-Sapiens language started to branch into different linguistic families only after many words were established as grammatical and derivational morphemes and compounding became common as a means to create new words. That process was very slow for long time, so Kenanidis (2013) estimates that there were no essential differences between languages until about 9000 BCE, marking the establishment of new ways of living after the last Ice Age, and that there was satisfactory mutual comprehension between different linguistic families until about the year 3000 BCE.

3.3 Grammar

The Proto-Sapiens grammar was so simple that the sporadic references in previous paragraphs have essentially described it. The prime importance of sound symbolism for the people of nature should be noted again before we further detail that the vowel “E” was felt as indicating the “yin” element, passivity, femininity etc., while “O” indicated the “yang” element, activeness, masculinity etc.; “A” was neutral or spiritual, indicating things conceived by the mind and emotions rather than with the physical senses. Some grammatical applications of this phonaesthesia were that the PIE verb stem had always “E” at least in the imperfective aspect, because the verb is a passive thing: something created by the subject. The Proto-Semitic subject of verbs was marked with -U (being also “yang” and active, as the older “O”), because the subject is the doer, the active principle; “A” was used for the plural of neuter nouns and for adverbial words already in the earliest documented forms of Greek.

A single consonant root that was often used, and so later became a grammatical marker, was the “S” indicating effort when added to some IE verbs, e.g., Greek streph- (= to turn), ster- (= to deprive), Latin “specio” (= to look); the correspondent Sanskrit verb “paśy-ami” lacks that added s-.

Another important, probably single consonant root was -CC- meaning “to do, to cause”: it is found as a prefix “k-” on Proto-Sino-Tibetan verbs, rendering them causative. The same root -CC- came to be the ergative case marker in numerous languages, like Sumerian and Basque. Even more importantly, the word for “I, myself” in many languages and whole linguistic families was based on this root (-CC-): PIE *eg’ō/(e)jo:/ originally meant “I do / I did” and not simply “I”; the pronoun “I” was simply conveyed by the -ō, as in all PIE active verbs (Note 5). Exactly corresponding to PIE *eg’ō/(e)jo:/ is the Proto-Austronesian *aku (= I, myself; in nominative), which is preserved until today as “ako/aku” in almost all Austronesian languages. Exactly that form “aku” is found in the nominative of the Akkadian pronoun for “I”, “an-aku” (there are similar forms in most other Semitic and Afro-Asiatic languages). A question remaining is the meaning of that “an-” in “anaku”, used for the oblique cases of the pronoun in Semitic and Afro-Asiatic. That “an-” (starting with a vowel or often as a mere “n-” or “ŋ-” in the more conservative languages as Sumerian and all Australian aboriginal languages) corresponds to the 1st person singular pronoun in many languages, e.g., Basque “ni”; Nahuatl “ni”; Sumerian “ŋae”; Korean “naega” (where we also see the -eg- corresponding to “ego”, “aku”, “an-aku”, mentioned previously). Of course, forms with “ŋ” are more original than those with “n”, because “ŋ” is a “difficult” sound, especially in the beginning of words. But it can also be observed that even more languages (including all Turkic, Manchu-Tungus, Mongolian, IE ones) have “m” in place of that ŋ/n for the oblique cases of the 1st person singular pronoun. So, was it “ŋ” or “m” the original root for “I” or “me”? No! Sumerian, the oldest recorded language, gives the clue: although “ŋae” is “I” in the main language, the feminine sociolect (called “Emesal”) has “mae” and generally turns the “ŋ” to “m” in other forms of the pronoun and in unrelated words too. The word for “woman” was “mi” in the main Sumerian language and the root *-m- for female is attested by many words and affixes in various linguistic families. Women could not use “ŋ” for “I”, because “ŋ” signified “man” or “male”, so they used “m” that meant “woman”: “this woman” = “me” in women’s speech, as “this man” = “me” in men’s speech. A Proto-Sapiens root *ŋar for “man, male”, expanded

from the mere “η”, is found in IE Sanskrit “nara” (= man) and Greek “anēr” (= man), also in Turkic “koç-ηar” (= male sheep, ram). So, this a- in Greek “anēr” is not due to a supposed laryngeal, but it was needed to ease the pronunciation of “η” in very ancient times (see Appendix F for the 2nd singular pronoun).

A very common grammatical morpheme found in early languages was a suffixed “i”, often lengthened for emphasis, becoming “j” in contact with vowels. This was the ending for derived adjectives (initially it was simply a kind of demonstrative pronoun or interjection: “of this kind”). This ending is common in Arabic adjectives, it is also used (as -i) in all Japanese descriptive adjectives, e.g., “nagai” (= long), “katai” (= hard, stiff), “kuroi” (= black), “shiroi” (= white) and so on. In Turkic, it is used in many adjectives derived by nouns, as in “tışı” (= female) from “tış” (= female genitals), “köti” (= thing of poor quality; coward person) from “köt” (= anus), “tejeri” (= heavenly) from *tejer (= heaven). In PIE, -j- is the usual suffix for adjectives derived from nouns (followed by the inflectional ending of the word), such as the Mycenaean Greek “wrine-j-os” (= made of leather) from “wrinos” (= skin). Also in ancient Greek, a long “i” is often added after any demonstrative pronoun to emphasize the point physically shown, e.g., “toutou” (= of that) – “toutouī” (= exactly of that one there). There is also the PIE relative pronoun [j + case-ending]; so we infer that there was a Proto-Sapiens root -J- used as a demonstrative pronoun (if that -J- did not evolve from the vowel “E”), which in later languages became a marker of adjectives.

3.4 Syntax

Modifiers to nouns (N) are basically adjectives (A), genitive case (G) and relative clauses (Rel), so today the word order of a language is shown as, e.g., AN, GN, RelN. All kinds of modifiers to nouns are equal to relative clauses. For example, when we say “a man’s jacket” we mean “a jacket which is worn by a man” or “which belongs to a man”; when we say “a green jacket”, we mean “a jacket which is green”, and so on. The Proto-Sapiens language did not know of modifiers to nouns; there was nothing that could be equaled to a relative sentence; all sentences were based on the subject-predicate pattern: there was a subject, expressed or (quite often) inferred, and a predicate to the subject. In fact, the subject-predicate pattern coincides with the topic-comment construction. What we call today “a verb” was a type of predicate, so it had to be placed at the end of the sentence. When a noun was combined with any type of predicate, the noun had to be placed first. The mind of primitive people had to follow a course from the more concrete to the more abstract; from the more familiar, to the more indirectly perceived. It is well known in modern psychology that verbs are abstract concepts in comparison to nouns. That is why children, until today, first learn to use concrete nouns, later descriptive adjectives, especially for colors, and even later, verbs.

As to words describing positions or placement in space or time, all those are only understood in relation to concrete nouns; i.e., to understand “on the table”, we must first conceive “table”; to understand “in the jar”, we must first conceive “the jar”; to understand “from the cave”, we must first conceive “the cave”, and so on. So, all those words for placement and relations were adverbs that became post-positions (Po) or suf-fixes. But even the adverbs were originally nouns. For example, take the PIE “pro” (= front, ahead): it was an adverb that previously was a noun, meaning “head”, and particularly “animal’s head”. This word is found as “buru” in Basque; “boro” in Embera (Columbia & Panama); “puś” in Chuvash; “baş” in common (z-)Turkic (all these meaning “head”; metaphorically “beginning”); “palil” in Sumerian (= a person going ahead, leader warrior, pioneer). The z-Turkic “ş” normally corresponds to “l” of the r-Turkic, including Chuvash, so the reconstructed Proto-Sapiens word was probably *pp(o),l – the retroflex /l/ nearing a rhotic sound. All this makes for a Subject-Object-Verb (SOV) or, more generally, V-final word order, which is still the most common among the language families of the world (Gell-Manna & Ruhlen, 2011).

More specifically, the first Homo Sapiens would not say “a green, men’s jacket was found on a seat; somebody lost it”, but literally “jacket, green, men’s, seat-upon found; it somebody lost”, which was interpreted as “a jacket is green; it is of men; there is a seat, and upon (that seat), it was found; that thing (the jacket) is what somebody lost”. This evolved to a syntax (V-final, NA, NG, NRel, Po, AdvV) that we know from the most ancient known languages: Sumerian and Elamite; also a few Australian aboriginal languages (notably Arrente), which represent a fossil of early civilization and language, use this word order, while almost all Australian aboriginal languages use this word order with the only difference being GN (genitive-noun). A huge body of especially Asian language families (Turkic, Manchu-Tungus, Korean, Japanese, Dravidian, IE, classical Nahuatl, and others) have gone one step further to AN (adjective-noun) and even RelN (relative-noun).

An objection may be raised because of the fact that a big group of languages, mainly Afro-Asiatic, Austronesian and (not so clearly) Mesoamerican languages like Mayan, show a V-initial (verb-initial) syntax. This however can easily be explained by the fact that “verb” was not a clearly defined grammatical category at the time that the Proto-Sapiens language began to branch out. So, a number of people that separated early from the rest of the human population did not use verbs at all; instead they used what would be best described as “participles” in modern

terminology. In fact, what is even today used in Austronesian languages is not really verbs, but various kinds of participles. For example, the most common verb-substitutes in Filipino languages are formed by the infixes -um- and -in-: these really form participles; in other words: nouns with a verbal sense. For example, the root “bili” (= buy) gives “b[um]ili” (= a thing that becomes bought) and “b[in]ili” (= a thing that has already been bought). So, they do start sentences with “bumili” and “binili”, but these sentences literally mean, for example, “a thing that becomes (or became at that time) bought, (was) my bicycle” and “a thing that is already bought, (is) my bicycle” (Note 6). Biblical Hebrew, mostly using a V-initial syntax (but sometimes also V-final), could make whole paragraphs without real verbs but only participles; a very modest example is the phrase “a voice of somebody shouting in the desert”, instead of “somebody shouts in the desert”. This type of syntax (using participles instead of verbs) accounts for languages that appear to use, or have really evolved to use, a verb-initial syntax.

4. Discussion & Conclusion

This description of the Proto-Sapiens language is of course approximate, and some parts or details of the reconstruction may be less accurate than other parts. The method that we call Qualitative Inquiry involves many years of serious study into as many languages as possible, especially early ones. Kenanidis (2013) reconstructed or suggested more than 300 roots of Proto-Sapiens, a big part of which was based on a reconstruction of the actual phonology and phonetic evolution of Sumerian. Bengtson & Ruhlen (1994) presented a similar list of 27 roots. Such lists are not likely to coincide unless in rare cases, mainly because the Proto-Sapiens language, having no adjectives but an extremely rich inventory of nouns, used many different terms for what seems to be the same thing; for example, a different root for flowing water (AQ), a different for discovered water (TTER), a different one for irrigating or needed water (WATT), another for clean water, another for still water, another for water in a hole, another for water meeting firm land (see Appendix G) and so on. Such a richness of vocabulary, unfamiliar to many modern nations, can be found today in languages often considered “primitive”, as many Austronesian ones are. Blust & Trussel (2010) give countless single words that need whole sentences (if not paragraphs) to be explained in English; a very modest example is “shiny speckles of oil floating on liquids or sauce”. Such terms give a clue to understand why the Proto-Sapiens language did not really need adjectives. Also, the meaning of some reconstructed roots can be eluding; for example, Ruhlen (1994c, p.103) reconstructs “putV” as “vulva”, but was this indeed the original meaning of the root? It is likely that the original meaning of “put-” was not vagina, but “something stinky or rotten”, judging from PIE *pudh (Greek puth- “to decay”; Sanskrit puti- “stinky”), and given the fact that the most common term of abuse among Australian aboriginals was “(oh you) vagina rotten (!)”. Also, Ruhlen (1994c) lists the words as if the vowels were a semantic part of the root, which is sharply in disagreement to ancient languages like Proto-Semitic, PIE, Proto-Kartvelian, and traces in other linguistic families, showing that only in later times the vowels were fixed in words. Furthermore, some of the terms he lists as Proto-languages’ forms are very questionable, e.g., “mala” is supposed to be Proto-Australian for “Arm-1”, while WD aboriginal languages have “mina” for “arm”, and that “mina” suggests a cognateness to Latin “manus”, so the real Proto-Australian is likely “mana” and not “mala”. And yet, it is not rational to list that “mala” with “Arm-1”, where most other linguistic families show “kan”, “kon” and such. He could put that “mala” perhaps with another list “Arm-3”, but he only has “Arm-1” and “Arm-2”, without explaining the difference between the two. Another example is “čuna” as “global term” for “nose”, but we have not found any evidence to justify a Proto-Language sound “č”. A listed Proto-word for nose is the Austric “ijun”, obviously based on Tagalog “ilon” and Bicol “iron”, but this ignores more conservative Filipino languages that have it as “duñó”, and the more likely reconstructed “global” term *dw-ŋ for “nose” could hardly match the supposed Eurasiatic “snā”. The space of this work does not allow for much more criticism on such methods of linguistic comparison; it will presently suffice to note that Ruhlen’s every list of proposed cognates contains incongruent words, e.g., “rt’s’q’a”, “nīru”, “namaw”, “gugu”, “nki”, and all these to justify a reconstructed global “ak’a” for “water”. This shows an unreliable application of a principle that “the majority prevails”, which is a type of quantitative method and not a qualitative one as used for our present work, which especially prescribes (see subsection 2.6) that only roots (and not words) should be compared, if we are to reliably reconstruct the features (subsections 3.2 and 3.4), including the vocabulary, of Proto-Sapiens. We do not oppose quantitative research methods, but extreme care should be taken in counting the right things in the right groups, and this care has been progressively neglected, while computers encourage researchers to process large quantities of data (List, 2017), which can easily be not the right data to process for each purpose.

It is concluded, on the other hand, that all those words presented as reconstructed “global / Proto-Human” words are witnesses; if even only one of so many witnesses is truthful, then the testimony is true that the whole population of Homo Sapiens used one and the same language. To the advocates of monogenesis, it is a soothing feeling to know that sound fundamentals have been laid for the reconstruction of the language that the ancestors of all the

present humanity used. Nobody can say that this reconstruction will be easy because we have to scrutinize far into the past, and we cannot be sure that controversy on the subject will be resolved soon, as this is an era of people not very much willing to “speak the same language”.

References

- Alexandre, F. (2014). Trees, Waves and Linkages: Models of Language Diversification. In C. Bowerman & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics* (pp. 161-189). London: Routledge.
- Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., & Rak, Y. (1989). A Middle Palaeolithic human hyoid bone. *Nature*, 338(6218), 758-760.
- Bengtson, J. D. (2007). *On fossil dinosaurs and fossil words*. Retrieved from <http://jdbengt.net/articles/fossilwords.pdf>
- Bengtson, J.D., & Ruhlen, M. (1994). Global Etymologies. In M. Ruhlen (ed.), *On the Origin of Languages: Studies in Linguistic Taxonomy* (pp. 277-336). Stanford: Stanford University Press.
- Blust, R. (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In E. Zeitoun & P. J. K. Li (eds.), *Selected Papers from the 8th International Conference on Austronesian Linguistics*. Taipei, Taiwan: Academica Sinica.
- Blust, R., & Trussel, S. (2010). *Austronesian Comparative Dictionary* (web edition). Retrieved from <http://www.trussel2.com/ACD/>
- Campbell, L. (2001). Beyond the Comparative Method. In B.J. Blake, K. Burridge & J. Taylor (eds.), *Historical Linguistics 2001* (13-17 August 2001). Melbourne: 15th International Conference on Historical Linguistics.
- Campbell, L. (2004). *Historical Linguistics: An Introduction* (2nd ed.). Cambridge: The MIT Press.
- Campbell, L., & Poser, W. J. (2008). *Language Classification: History and Method*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486906>
- Christiansen, M. H., & Kirby, S. (2003). Language evolution: the hardest problem in science? In M.H. Christiansen & S. Kirby (eds.), *Language evolution* (pp. 77-93). Oxford & New York: Oxford University Press.
- Clauson, G. (1972). *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford: At The Clarendon Press.
- Crowley, T. (1992). *An Introduction to Historical Linguistics* (2nd ed.). Auckland: Oxford University Press.
- Dixon, R. M. W., & Aikhenvald, A. Y. (2001). *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press.
- Dolgopolsky, A. (1964). Gipoteza drevnejšego rodstva jazykovyx semei severnoj Eurasii sverojatnostnoj točki zrenija. *Voprosy Jazykoznanija*, 2, 53-63 (translated into English by V.V. Shevoroshkin & T.L. Markey [eds.], *Typology, Relationship and Time*, 1986 [pp. 27-50]. Mich.: Ann Arbor).
- Durie, M., & Ross, M. (1996). *The comparative method reviewed: Regularity and irregularity in language change*. New York: Oxford University Press.
- Embleton, S. (1992). Historical Linguistics: Mathematical concepts. In W. Bright (ed.), *International Encyclopedia of Linguistics*. New York: Oxford University Press.
- Endicott, P., Metspalu, M., & Kivisild, T. (2007). Genetic evidence on modern human dispersals in South Asia: Y chromosome and mitochondrial DNA perspectives: The world through the eyes of two haploid genomes. In M.D. Petraglia & B. Allchin (eds.), *The Evolution and History of Human Populations in South Asia* (pp. 229-244). Netherlands: Springer.
- Fleming, H. C. (2003). Newsletter of the Association for the Study of Language in Prehistory (August 2003). *Long Ranger*, 34(b). Retrieved from <http://www.aslip.org/LR/MTLR-34b.htm>
- Fox, A. (1995). *Linguistic Reconstruction: An Introduction to Theory and Method*. New York: Oxford University Press.
- Gell-Manna, M., & Ruhlen, M. (2011). The origin and evolution of word order. *PNAS*, 108(42), 17290-17295. <https://doi.org/10.1073/pnas.1113716108>
- Givón, T. (2000). Internal reconstruction: As method, as theory. In Spike Gildea (ed.), *Reconstructing grammar: comparative linguistics and grammaticalization* (pp. 107-160). Amsterdam/Philadelphia: John Benjamins.
- Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-

- European origin. *Nature*, 426, 435-439. <https://doi.org/10.1038/nature02029>
- Greenberg, J. H. (1957). *Essays in Linguistics*. Chicago: University of Chicago Press.
- Greenberg, J. H. (2005). *Genetic Linguistics: Essays on Theory and Method* (edited by William Croft). Oxford: Oxford University Press.
- Gudschinsky, S. C. (1956). The ABC'S of Lexicostatistics (Glottochronology). *Word*, 12(2), 175-210. <https://doi.org/10.1080/00437956.1956.11659599>
- Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *PNAS*, published ahead of print January 29, 2018. <https://doi.org/10.1073/pnas.1713975115>
- Heggarty, P., Maguire, W., & McMahon, A. (2010). Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. *Philosophical Transactions of the Royal Society, B*, 365, 3829-3843. <https://doi.org/10.1098/rstb.2010.0099>
- Hock, H. H., & Joseph, B. D. (1996). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- Hockett, C. F. (1960). The Origin of Speech. *Scientific American*, 203(3), 88-96. <https://doi.org/10.1038/scientificamerican0960-88>
- Hymes, D. (1960). Lexicostatistics so far. *Current Anthropology*, 1(1), 3-44. <https://doi.org/10.1086/200074>
- Jackson, J. P., & Weidman, N. M. (2005). *Race, Racism, and science: social impact and interaction*. New Brunswick, NJ: Rutgers University Press.
- Kalgren, B. (1923). *Analytic dictionary of Chinese and Sino-Japanese*. New York: Dover publications, 1974 [Librairie Orientaliste Paul Geuthner, Paris 1923].
- Kenanidis, I. (1992). *Eteokreetes Megaleetores*. Athens: National Library of Greece (in Greek).
- Kenanidis, I. (2013). *Historical and Linguistic Studies: Homosapiens.htm*. Kavala: E. P. Lazidou (in Greek).
- Kenanidis, I., & Papakitsos, E.C. (2013). Yet another suggestion about the origins of the Sumerian language. *International Journal of Linguistics*, 5(5), 30-44. <http://dx.doi.org/10.5296/ijl.v5i5.4107>
- Kiselman, C. (2008). Esperanto: Its origins and early history. In A. Pelczar (ed.), *Prace Komisji Spraw Europejskich PAU* (vol. II, pp. 39-56). Krakow: Polska Akademia Umieje.
- Koerner, E. F. K. (1999). *Linguistic historiography: projects & prospects* (Amsterdam studies in the theory and history of linguistic science; Ser. 3: Studies in the history of the language sciences). Amsterdam: J. Benjamins. <https://doi.org/10.1075/sihols.92>
- Kozyrski, W. H., & Malovichko, A. V. (2015). To Human Language Origins. *Journal of Language, Linguistics and Literature*, 1(3), 86-93. Retrieved from <http://www.aiscience.org/journal/j31>
- Labov, W. (2007). Transmission and diffusion. *Language*, 83, 344-387. <https://doi.org/10.1353/lan.2007.0082>
- Lehmann, W. P. (1993). *Theoretical Bases of Indo-European Linguistics*. London: Routledge.
- Lieberman, P. (2007). The evolution of human speech. *Current Anthropology*, 48(1), 39-66. <https://doi.org/10.1086/509092>
- List, J. M. (2017). *A web-based interactive tool for creating, inspecting, editing, and publishing etymological datasets*. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, System Demonstrations (pp. 9-12). <https://doi.org/10.18653/v1/E17-3003>
- López, S., Dorp, van L., & Hellenthal, G. (2015). Human Dispersal Out of Africa: A Lasting Debate. *Evolutionary Bioinformatics*, 11(s2), 57-68. <https://doi.org/10.4137/EBo.s33489>
- Lyovin, A. V. (1997). *An Introduction to the Languages of the World*. New York: Oxford University Press.
- McMahon, A., & McMahon, R. (2005). *Language Classification by Numbers*. Oxford: Oxford University Press.
- Metspalu, M., Kivisild, T., Bandelt, H. J., Richards, M., & Villems, R. (2006). The Pioneer Settlement of Modern Humans in Asia. In H.-J. Bandelt, V. Macaulay & M. Richards (eds.), *Nucleic Acids and Molecular Biology, Vol. 18: Human Mitochondrial DNA and the Evolution of Homo sapiens* (pp. 181-199). Berlin & Heidelberg: Springer-Verlag.
- Pagel, M., Atkinson, Q. D., Calude, A. S., & Meade, A. (2013). Ultraconserved words point to deep language

- ancestry across Eurasia. *PNAS*, 110(21), 8471-8476. <https://doi.org/10.1073/pnas.1218726110>
- Renfrew, C., McMahon, A., & Trask, L. (eds.) (2000). *Time Depth in Historical Linguistics*. Cambridge: The McDonald Institute for Archaeological Research.
- Rowe, B. M., & Levine, D. P. (2015). *A Concise Introduction to Linguistics*. New York: Routledge.
- Ruhlen, M. (1994a). The Origin of Language: Retrospective and Prospective. In M. Ruhlen (ed.), *The Origin of Languages: Studies in Linguistic Taxonomy* (pp. 261-276). Stanford: Stanford University Press.
- Ruhlen, M. (1994b). *On the Origin of Languages: Studies in Linguistic Taxonomy*. Stanford: Stanford University Press.
- Ruhlen, M. (1994c). *The Origin of Language: Tracing the Evolution of the Mother Tongue*. New York: John Wiley and Sons.
- Ruhlen, M. (2005). Taxonomy, Typology and Historical Linguistics. In J.W. Minnet & W.S.-Y. Wang (eds.), *Language Acquisition, Change and Emergence* (pp. 341-368). Hong Kong: City University of Hong Kong Press.
- Schleicher, A. (1874). *A Compendium of the Comparative Grammar of the Indo-European, Sanskrit, Greek and Latin Languages* (translated into English by H. Bendall). London: Trübner and Co.
- Schwink, F. W. (1994). *Linguistic Typology, Universality and the Realism of Reconstruction*. Washington: Institute for the Study of Man.
- Starostin, S. (2002). *Methodology of Long-Range Comparison*. Retrieved from <http://starling.rinet.ru/Texts/method.pdf> (accessed on March 27th, 2018).
- Starostin, S. A., & Bronnikov, Y. (1998-2009). *Languages of the World Etymological Database*. Part of the Tower of Babel - Evolution of Human Language Project. Retrieved from <http://ehl.santafe.edu/main.html> (accessed March 4, 2018).
- Stix, G. (2008). The Migration History of Humans: DNA Study Traces Human Origins Across the Continents. *Scientific American*, 299(1), 56-63. <https://doi.org/10.1038/scientificamerican1108-56>
- Stringer, C. (2011). *The Origin of Our Species*. London: Penguin.
- Swadesh, M. (1955). Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics*, 21, 121-137. <https://doi.org/10.1086/464321>
- Swadesh, M. (1972). *The origin and diversification of languages*. London: Routledge & Kegan Paul.
- Szemerényi, O. J. L. (1996). *Introduction to Indo-European Linguistics* (4th ed.). Oxford: Oxford University Press.
- Ulbæk, I. (1998). The origin of language and cognition. In J.R Hurford, M. Studdert-Kennedy & C. Knight (eds.), *Approaches to the evolution of language: social and cognitive base* (pp. 30-43). Cambridge & New York: Cambridge University Press.
- Yi-chün Lu, A. (1998). *Phonetic motivation: a study of the relationship between form and meaning*. München: Hieronymus.
- Zhivotovsky, L. A., Rosenberg, N. A., & Feldman, M. W. (2003). Features of Evolution and Expansion of Modern Humans, Inferred from Genomewide Microsatellite Markers. *American Journal of Human Genetics*, 72(5), 1171-1186. <https://doi.org/10.1086/375120>.
- Zorc, R. D. (1995). A Glossary of Austronesian Reconstructions. In D. T. Tryon (ed.), *Comparative Austronesian dictionary: An Introduction to Austronesian Studies* (Part 1: Fascicle 2, pp. 1106-1197). New York: Mouton de Gruyter.

Notes

Note 1. Namely: thou, I, not, that, we, to give, who, this, what, man/male, ye, old, mother, to hear, hand, fire, to pull, black, to flow, bark, ashes, to spit, worm.

Note 2. Namely: Proto-Dravidian, Proto-Kartvelian, Proto-Uralic, Proto-Indo-European, Proto-Altaic, Proto-Chukchi-Kamchatkan and Proto-Inuit-Yupik.

Note 3. For example, the words for “body” in Italian/Portuguese: “corpo”; Spanish: “cuerpo”; French: “corps”.

Note 4. This can explain why it is very difficult today to find cognate roots between Semitic and other Afro-Asiatic

languages: because the original roots were treated differently in each group.

Note 5. By the way, the 1st person singular pronoun is “u” in Cuneiform Elamite, a very ancient language.

Note 6. Although European language speaking people find it more convenient to translate “then I bought my bicycle” and “I have already bought my bicycle”.

Appendix A

Mistaken Cognates

All linguists know the example of Latin “habeo” and German “habe” which have the same meaning (= I have) but are neither cognates nor loanwords. Could such a fragile sound like /h/ remain unchanged for so many centuries? Wherever a /h/ is found in IE languages, it originates from something else with scarcely any exception beyond Hittite and Luwian. Besides, as far as we know, no old language had a verb meaning “I have”; e.g., the Ancient Greek “ekh-”, originates from the PIE root “sek’h-”, meaning “to hold”.

Another interesting example is “mata” (= eye) in many Austronesian languages, which some people compare to Modern Greek “mati” (= eye)! Of course, this similarity is fortuitous, because “mati” comes from a totally dissimilar root and Austronesian “mata” is from the reconstructed Proto-Austronesian “maCa”, i.e., /mat^ha/, which indeed is cognate to the IE Greek root math- (today meaning “to learn”, but the ancient meaning was “to notice, realize, understand”), because it is usually through the eye that people notice things. This is a good example to show that word similarity means nothing, but root similarity is enlightening.

Appendix B

Individual Phonemes Standing as Cores of Lexical Roots

In etymological dictionaries of IE languages it has been noticed that old IE words about mental functions start with m-, such as Greek “menos” (= mind, mental power), “mētis” (= plan), med- (= to plan, care of), “metron” (= measurement) etc. It has also been noticed that ancient IE words starting with “p” convey a sense of hardness or steadiness, e.g., PIE pod- (= legs), “pesos” (= penis), Greek “petra” (= rock), “pagos” (= rocky height), Latin “palus” (= a set up pole) etc. Kenanidis (2013) notices that the PIE bh- (corresponding to similar phonemes in other linguistic families) always has the meaning of “appearance”: “bh(e)w” (= to come into existence, sprout: also compare to WD aboriginal /puŋu/ “plants”, Filipino punó “tree”); “bhen” (Greek phain-, Sanskrit bhan-: to shine, shed light on); but also the Greek phōt- “an adult (respected) man”, compare to Japanese “hito” (*pito: person) and WD aboriginal “wati” (/wōti/: adult man; the important person in aboriginal society); even the latter has a meaning related to “appearance”: the person who has appeared in public and made an impression, being in privacy before. There are many other PIE roots starting with bh- and it can be observed that all those have a meaning related to “appearing (in an impressive way)”.

Appendix C

Phonemic Change

The case of classical Greek “phōs” (= light) and “pūr” (= fire) is really enlightening. These two are of PIE origin, “phōs” from “phawos” from PIE */b^həxwos/ and “pūr” from PIE /pəxwor/, which is attested although distorted by the Cuneiform script as “paḥur” in Hittite. /b^həxwos/ and /pəxwor/ resemble each other (both in meaning and form) in a way that reveals the same Proto-Sapiens word as evolved in two different language families. It was the Turkic and closely related languages (also called “Altaic languages”) that since very ancient times (since the 6th millennium BCE, judging from Sumerian, which can be classified with the r-Turkic/Altaic languages: Kenanidis & Papakitsos, 2013) lost the distinction between aspirated and non-aspirated consonants, while IE languages preserved this distinction that existed in Proto-Sapiens. The ending –os // –or existed in both forms in all ancient languages, notably in Turkic and in IE (also in Chinese as 子 “zi”); it was a kind of article that also served as a nominalizing suffix; it is found with both –s and –r within IE too, e.g., Latin “honos”, later “honor”. This all means that /b^həxwos/ was the original PIE word from a Proto-Sapiens root, while /pəxwor/ was borrowed into PIE from some ancient Turkic (or some closely related, Altaic) language. The Qualitative Inquiry then asks: is this confirmed by any Turkic or Altaic language preserving a form of that /pəxwor/? Absolutely yes: firstly, the word is found in Sumerian as “ḫiṣ-bar”, meaning “sacred fire”, which was pronounced /ḫixβar/ or /ḫixwar/ (if we clear the distortion that the Cuneiform script always causes). That /ḫixβar/ or /ḫixwar/ clearly comes from /pixwar/ from older /pəxwor/, according to perfectly well known phonetic changes in Sumerian, especially the very common change of “p” to “ḫ” (Kenanidis & Papakitsos, 2013: phonetic rule 5.0.15). Modern Japanese preserves the following forms of /pəxwor/ or its root /pəxw/:

火, 灯, 燈, 陽, 日 (read “hi” from */pix/ from */pəxw/) = bright fire, figuratively the sun;

陽か, 燈か, 灯か, 火か, 灯か, 日か (read “hika” from older */pixar/ from */pəxwor/) = the effulgence of fire or of the sun;

光 光り (read “hikari”, from */pixari/ from */pəxwori/) = bright, shining (like fire);

光る 晃 玉 (read “hikaru”, from */pixaru/ from */pəxworə/) = a verb meaning “to shine” (like a fire, or the sun, or a gem reflecting the light).

Although it is not the specific purpose of this work to show that Japanese is an Altaic/Turkic language of the eastern (z-Altaic/Turkic) group, we cite a few cognates between Japanese (J.) and z-Turkic (T.): J. -ra, T. -lar = a plural suffix; J. sake, T. çakır = fermented juice, wine; J. koshi = waist, T. kuşan- = “to girdle, wear or carry on the waist” (that koshi proves that J. is of the z-Altaic/Turkic group, because the r-Altaic/Turkic languages have “l” in place of “ş”).

All etymological dictionaries of IE languages and Turkic state that Turkic “öküz” is an early loanword from some IE language; this is a case of academic prejudice. The word “öküz” has a totally typical Turkic form and it is common from ancient times until today in almost all Turkic languages, including the r-Turkic languages, from which it was borrowed in ancient times into Mongolian as “(h)üker”. On the other hand, among IE languages the word has no convincing etymology as from any root and it is found only in some IE language subgroups, which had a contact with Asia; it is not found in Anatolian, Greek, or Italic languages, for example. Clauson (1972) wrote: “The (Tocharian) A form has survived only as *o[gap]* in a much damaged text but can be restored fairly confidently as *okäs*” (which hypothetical “okäs” is supposed to be the origin of the Turkic “öküz”); this is how “fairly confidently” the theory was formed! In fact the word existed in Turkic even before it split into the r- and the z-varieties; it is documented in Sumerian as “ug” or “uku₄” (i.e., “ökü”), where it is supposed to mean a lion or some ferocious feline animal. In fact, the word meant “the (most typical) strong animal”, which for the ancient Turks was the ox, used for its strength and being the symbol of strength (or even of a strong state) in the “Irk Bitig”, an ancient Turkic book of oracular parables. We know that the word (Ochse, ox etc.) cannot be an original IE word, because the Proto-Sapiens (PS) root that gave “öküz” in Turkic is the same Proto-Sapiens root that gave a very famous PIE word: “ek’wos” (/ecwos/) = horse; because the root (*H-W-C) in Proto-Sapiens meant “the strong animal”; “the strong animal” for the Indo-Europeans was “ek’wos”, the horse; the domesticated ox was a later cultural innovation that they received from the ancient Turkic (or “Altaic”) nation.

Appendix D

Comparisons between Different Linguistic Families

Those who think that Austronesian had an uvular trill (rhotic) are based on some words, which appear with “r” in some languages where other languages have “g”. For example, tail is “ikór” in Malay and “ikóg” in Bicol; rice cooked or prepared for cooking is “bägás” and similar other forms with “g” in most languages, but some languages have “beras” etc. with “r” in place of “g” (fewer languages have “h” there and “beyas” is found in one language); so most scholars reconstruct the original sound as */r/. However, in cases like “bägás”, the “r” in place of “g” can easily be explained as a secondary evolution from the original “g” (possibly through an intermediate */r/ or /y/ that was not preserved because these are difficult sounds; /y/ in this position is known from one language only: Manobo). The hypothesis of the original “g” is confirmed when we compare to languages of other families: that “bägás” is cognate to PIE bhag’- “food” (hence Greek phag- “food”; Phrygian “bekos”: bread), but also to words of other languages that were borrowed into IE languages, such as Sanskrit “bija” (/bi:jä/) = grains and probably Greek “pīs-on” (from */p^he:c:/) = peas and “phakē” (= lentils), borrowed from different languages. In the case of “ikór” and “ikóg” (= tail), there is an obvious similarity and probably cognateness to the Greek “kerkos” (= tail), Turkic “kürk” (= fur); thus “ikór” and “ikóg” probably come from *ikork (root *c-rc with a prefixed “i”), so again it was nothing like a */r/. Accordingly, comparisons between different linguistic families can help in determining a root within the same linguistic family too.

Appendix E

Semitic Roots Formation

The Semitic root for “sky” is Š-M, originating from Proto-Sapiens TTJ-D (= the sky as animate and therefore sacred). The root TTJ-D is also the root of Sumerian “dejer”(= God, sky) and Turkic “tejeri” (= God, sky), Chinese “tiān” (天 = sky, God), Polynesian “tangaloa” (= God), Proto-Austronesian *“laŋiC” (= sky; individual Austronesian languages have it usually with l-, but d-, r-, g-, x- and e- are also found) and PIE djēn- being the origin of the genitive “Ζηνός”, dative “Ζηνί” and accusative “Ζηνα” (the sky God Zeus) in Greek; PIE “dj(e)n” has given many derivative words in Sanskrit too, e.g., “dina” (= the light of sky). Then, that Semitic root ŠM was

expanded with a Š as Š-M-Š (= sun).

While TTJ-D was the Proto-Sapiens root for the sky as animate and sacred, the Proto-Sapiens root for the sky as simply a natural phenomenon was Q-C: this is attested by Turkic “kök” (= sky; blue color), Elamite “kek” (= sky) and Sanskrit “ākāśa” (= sky, aether; this “ākāśa” must be from an ancient borrowing into Sanskrit, where a /c/ turned to “ś”, according to Satem-IE phonotactics, and the preposition ā- was added). Then the Proto-Sapiens Q-C was expanded in Semitic as “kakab” (Akkadian), “kokab” (כֹּכַב Hebrew), “kukab” (كوكب Arabic), with the addition of -b.

Appendix F

Origins and Evolution of Pronouns

As explained in subsection 3.2 and 3.3, the only original pronouns were single vowels, “O” for the first person; the pronoun for the second person was “A”, for the third person: “E”. However, in descendant languages we find many other pronouns, one of the commonest can be reconstructed as TT-W with the obvious meaning “thou, you”: this is found in many diverse linguistic families, for example PIE and Turkic, including Sumerian, sometimes a TTW- having been spirantized into s- or z-. Although the original root TT-W and its later forms were used for “thou”, this root originally meant “a person” and not “thou”. They simply used “person” in vocative, just like Greek “ánthrōpe”, to mean “you”. Some language families still kept the root TT-W for “person, human being”, as Proto-Austronesian “tawo” and Proto-Bantu “ntu”. The extensive group of Afro-Asiatic languages usually forms the 2nd singular pronoun by joining the old “A” (“you”) to the old root for “person”, thus making it “atta” or “anta”, while the latter is also cognate to the Japanese “anata” (“you”).

Appendix G

Water-related Words

“QAR” was a Proto-Sapiens word describing the water that meets solid land and thereby the water rises. The root is found in Sumerian as “kar”, meaning the river port of every Sumerian city, named “kar” because it is where water meets the land. This Sumerian “kar” is the real origin of German “Kai” and English (through French) “quay” (a Sumerian word-final -r when not followed by a vowel could not be pronounced unless as -/j/). “ka:r-” is also a Turkic verb meaning for water “to pile up behind an obstacle” or for tide “to rise”, sometimes also used for canals overflowing (these definitions from: Clauson, 1972): the original meaning of the root was very well preserved in this Turkic verb (“water meets solid land”). The same root gave the WD aboriginal “karu”, meaning the temporary rivers in the (semi)arid areas, because the “karu” is the water coming to meet the land. Greek has this root, inherited from PIE, as “kor-os”, which means “to reach fullness”, to get “full to the brim”. The meaning of this root is also well preserved in Chinese “kâu” (皋 or 臯); it is a very old word because it is written by an original picture character and not with any phonetic element or even a logical compound character; it means “river bank; marsh; high, eminent” (Kalgren, 1923), the meanings “high, eminent” taken from the water that rises when it meets an obstacle. No Chinese words end with “l” or “r”, because every -l has turned to -w (written “u” or “o”), and most of the “r” turned to “l” since old times in Chinese. Another example of old Chinese “-l” to “-w” is 道 (Middle Chinese “d’âu”: Kalgren, 1923) meaning “route”, corresponding to Proto-Austronesian “dalan” (route), Turkic “yol” from */ðol/ from *dal (= route), Sumerian “sila” (/θe|ə/: route) with a much less common “tilla” (/t̪e|ə/: road).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).