Evolution of Eurasian and African Family Systems, Cross-Cultural Research, Comparative Linguistics, and Deep History*

Andrey Korotayev  
*National Research University Higher School of Economics, Moscow; Institute of Oriental Studies, Russian Academy of Sciences, Moscow

Svetlana Borinskaya  
Vavilov Institute of General Genetics and Peter the Great Museum of Anthropology and Ethnography, Russian Academy of Sciences, Moscow and St Petersburg

George Starostin  
National Research University Higher School of Economics, Moscow; Russian State University for the Humanities, Moscow

Kira Meshcherina  
National Research University Higher School of Economics, Moscow

ABSTRACT

Currently, the main source for the reconstruction of the most ancient history of humankind is archeology, which almost by definition makes it possible to restore only just a few elements of the most ancient human culture (naturally, almost exclusively – material culture). A mere introduction of comparative linguistic data makes it possible to significantly refine our reconstruction of a respective culture. If a certain linguistic Urheimat may be localized in space and in time within the area and lifespan of a certain archeological culture, this suggests that we may have an idea of the lan-
guage spoken by respective population, as the application of comparative linguistic methods allows us to reconstruct the vocabulary of the carriers of the respective protolanguage, that makes it possible to identify a set of terms denoting the realities of family organization, political attitudes, beliefs, etc. A very important part of the reconstructed vocabulary is constituted by the kinship terminology. As is well known (and as is demonstrated in this article again), the kinship terminology displays rather strong correlations with respective types of kinship organization, which could allow to reconstruct important features of clan and family structure of the respective populations. This reconstruction can be further verified by using archaeological and genetic data. It is demonstrated that the papers presented at the International Workshop ‘Murdock and Goody Re-visited: (Pre)history and evolution of Eurasian and African family systems’ that was organized in April 2015 by the Max Planck Institute for Social Anthropology suggest that we are close to having all the necessary ingredients to undertake such a program of a deep historical reconstruction.

INTRODUCTION

By now we observe the emergence of an idea that a significant breakthrough in the reconstruction of the most ancient history of humankind in an unusually full detail can be achieved through the synthesis of the methodologies of sociocultural anthropology, comparative linguistics, cross-cultural research, genetics, archeology, comparative folklore studies, and so on (see, e.g., Ember et al. 2006; Korotayev and Kazankov 2000; Jones 2003; Korotayev 2006; Korotayev et al. 2006; Smail 2007; Gray, Drummond, and Greenhill 2009; Fortunato 2011; Gray, Atkinson, and Greenhill 2011; Shenk and Mattison 2011; Bouckaert et al. 2012; Bellwood 2014; Holman et al. 2015; Johnson and Paul 2016).

Currently, the main source for the reconstruction of the most ancient history of humankind is archeology (see, e.g., Ember, Ember, and Peregrine 2015), which almost by definition makes it possible to restore only just a few elements of the most ancient human culture (naturally, almost exclusively – material culture). For example, the history of Syria and Palestine between 12,500 and 9,500 BCE is reconstructed, first of all, on the basis of archaeological data on the Natufian culture (see, e.g., Bar-Yosef and Valla 1992; Akkermans and Schwartz 2003; Simmons 2007, etc.). Note that in this case (as well as in all the other similar cases), the archaeological data do not allow us to reconstruct in any reasonable detail the social organization of the ‘Natufians,’ nor their system of beliefs, not to mention things like music and rituals. A mere introduction of comparative linguistic data makes it possible to significantly refine our reconstruction of this culture. When we turn to comparative linguistic data, we find out that Afroasiatic Urheimat may be localized in space in time with the area and lifespan of the Natufian culture (Militarev 2000; Starostin 2000; Ember et al. 2006).
The application of comparative linguistic methods allows us to reconstruct the vocabulary of the carriers of the proto-Afroasiatic language, that makes it possible to identify a set of terms denoting the realities of family organization, political attitudes, beliefs, etc. (Militarev 1990, 2000, 2002, 2004). Of course, as soon as we have some idea about the language spoken by those people who lived in a certain archaeological site, this allows us to reconstruct their culture in a much more detail than when we have at our disposal archaeological materials only. A very important part of the reconstructed vocabulary is constituted by the kinship terminology. As is well known, the kinship terminology displays rather strong correlations with respective types of kinship organization, which could allow us to reconstruct important features of social organization of the respective populations on the basis of quantitative cross-cultural analyses.


For an illustration, we will present below some results of our own tests with Murdock’s Ethnographic Atlas database. Though such patterns like the correlation between the Omaha kinship terminology and patrilineal descent groups, or the correlation between the Crow kinship terminology and matrilineal descent groups were established rather long ago (see, e.g., White 1939; Murdock 1949, 1968; Textor 1967; Goody 1970), it might be difficult to imagine this, but no formal tests of these correlations have been performed since Textor (1967) who did not have at his disposal a full version of Murdock’s Ethnographic Atlas (that was far from being finished by the time of Textor’s enterprise) and who used rather archaic statistical techniques; hence, it appears appropriate to re-test such correlations using a full version of the Ethnographic Atlas database as well as more advanced statistical techniques. This will allow us to find in a somehow accurate way the answer to the following question that appear immensely important in the framework of the present paper: if we know the kinship terminology of a certain human culture, how well could this terminology predict the clan and family organization of a group belonging to this culture?

In fact, the degree to which the knowledge of the kinship terminology could predict the knowledge of the clan and family social organization turns out to be rather high.

For example, if the reconstructed kinship terminology for a certain protolanguage turns out to be non-bifurcate (either Hawaiian, or Inuit) we would have certain grounds to hypothesize that its speakers lacked unilinear descent groups (see Table 1).
Table 1

Correlation between the Presence of Non-Bifurcate Kinship Terminology and the Absence of Unilineal Descent Groups

<table>
<thead>
<tr>
<th>Kin terms</th>
<th>0 = absent</th>
<th>1 = present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = non-bifurcate</td>
<td>70.6 % (72)</td>
<td>29.4 % (30)</td>
<td>102</td>
</tr>
<tr>
<td>1 = bifurcate</td>
<td>31.1 % (130)</td>
<td>68.9 % (288)</td>
<td>418</td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>318</td>
<td>520</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. \( p << .0001 \) (by Fisher's exact test), \( \phi = +.322; \gamma = + 0.683, p << .0001. \)

If it turns out to be specifically Inuit, we could suppose this with an even higher degree of confidence (see Table 2).

Table 2

Correlation between the Presence of the Inuit Kinship Terminology and the Absence of Unilineal Descent Groups

<table>
<thead>
<tr>
<th>Inuit kin terms</th>
<th>0 = absent</th>
<th>1 = present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = absent</td>
<td>28.3 % (325)</td>
<td>71.7 % (824)</td>
<td>1149</td>
</tr>
<tr>
<td>1 = present</td>
<td>82.2 % (83)</td>
<td>17.8 % (18)</td>
<td>101</td>
</tr>
<tr>
<td>Total</td>
<td>408</td>
<td>842</td>
<td>1250</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. \( p << .0001 \) (by Fisher's exact test), \( \phi = -.313; \gamma = - 0.842, p << .0001. \)

The Crow kinship terminology is a rather strong predictor of the presence of matrilocal / avunculocal residence or matrilocal / avunculocal alternative (see Table 3).
Table 3

Correlation between the Presence of the Crow Kinship Terminology and the Presence of Matrilocal / Avunculocal Residence or Matrilocal / Avunculocal Alternative

<table>
<thead>
<tr>
<th>Crow kin terms</th>
<th>Matrilocal / avunculocal residence or matrilocal / avunculocal alternative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = absent</td>
<td>70.6 % (608)</td>
<td>861</td>
</tr>
<tr>
<td>1 = present</td>
<td>24.6 % (14)</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td>622</td>
<td>296</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. $p << .0001$ (by Fisher's exact test), $\phi = +.238; \gamma = + 0.761, p << .0001.$

It is an even stronger predictor of the presence of matrilineal descent groups (see Table 4).

Table 4

Correlation between the Presence of the Crow Kinship Terminology and the Presence of Matrilineal Descent Groups

<table>
<thead>
<tr>
<th>Crow kin terms</th>
<th>Matrilineal descent groups</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = absent</td>
<td>85.4 % (746)</td>
<td>874</td>
</tr>
<tr>
<td>1 = present</td>
<td>14.3 % (8)</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>754</td>
<td>176</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. $p << .0001$ (by Fisher's exact test), $\phi = +.432; \gamma = + 0.944, p << .0001.$

On the other hand, the Omaha kinship terminology is a rather strong predictor of the presence of viri-/patrilocal residence (see Table 5).
**Table 5**

Correlation between the Presence of the Omaha Kinship Terminology and the Presence of Viri-/Patrilocal Residence

<table>
<thead>
<tr>
<th>Viri-/Patrilocal Residence</th>
<th>Omaha kin terms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
</tr>
<tr>
<td>0 = absent</td>
<td>36.1 % (304)</td>
</tr>
<tr>
<td>1 = present</td>
<td>9.1 % (7)</td>
</tr>
<tr>
<td>Total</td>
<td>311</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. $p < .0001$ (by Fisher's exact test), $\phi = .158; \gamma = + 0.70, p << .0001$.

It is an even stronger predictor of the presence of patrilineal descent groups (see Table 6).

**Table 6**

Correlation between the Presence of the Omaha Kinship Terminology and the Presence of Patrilineal Descent Groups

<table>
<thead>
<tr>
<th>Patrilineal descent groups</th>
<th>Omaha kin terms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
</tr>
<tr>
<td>0 = absent</td>
<td>57.2% (487)</td>
</tr>
<tr>
<td>1 = present</td>
<td>7.7% (6)</td>
</tr>
<tr>
<td>Total</td>
<td>493</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. $p << .0001$ (by Fisher's exact test), $\phi = +.275; \gamma = + 0.882, p << .0001$.

However, the presence of the Omaha kinship terminology is an even stronger predictor of the presence of patrilineal inheritance rule for real property; in fact, its predictive force in this case turns out to be rather close to the maximum value (see Table 7).
Table 7

Correlation between the Presence of the Omaha Kinship Terminology and the Presence of Patrilineal Inheritance Rule for Real Property

<table>
<thead>
<tr>
<th>Omaha kin terms</th>
<th>Patrilineal inheritance rule for real property</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
<td>1 = present</td>
</tr>
<tr>
<td>0 = absent</td>
<td>32.5 % (119)</td>
<td>67.5 % (247)</td>
</tr>
<tr>
<td>1 = present</td>
<td>2.9 % (1)</td>
<td>97.1 % (34)</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>281</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. $p < .0001$ (by Fisher's exact test), $\phi = +.183; \gamma = + 0.885, p << .0001$.

Below we will provide an example how the reconstruction of the kinship terminology of a certain protolanguage can provide some inferences about social kin and family organization of the speakers of that protolanguage. We will use the most profoundly reconstructed protolanguage (the proto-Indo-European) as an example (for the linguistic reconstruction of proto-Indo-European see, e.g., Benveniste 1969, 1973; Beekes 1995; Clackson 2007; Gamkrelidze and Ivanov 1995; Fortson 2004; Mallory and Adams 2006; Meier-Brügger 2003; Pokorny 2005).

For example, judging by reflexes in descendent languages the proto-Indo-European *aw- appears to have denoted both ‘Father’s Father (FaFa)’ (cf. Latin patris mei pater aus meus est ‘My father’s father is my ausus’ [Benveniste 1969, I: 226]) and ‘Mother’s Brother (MoBr)’, whereas the proto-Indo-European *nepo-t-er- seems to have denoted both ‘Sister’s Son (SiSo)’ and ‘Son’s Son (SoSo)’. As a result we seem to confront in proto-Indo-European a rather peculiar combination of kinship terms: ‘FaFa = MoBr’ and ‘SiSo = SoSo’ (see, e.g., Benveniste 1973; Gamkrelidze, Ivanov 1995).

An important point is that it seems difficult to imagine any other context within which each couple of those rather peculiar relatives could be denoted with one term, but matrilineal patrilateral cross-cousin one.

Indeed, with systematic patrilateral cross-cousin marriage this group of relatives would be arranged in the following way (see Fig. 1 below).
Fig. 1. A reconstruction of the proto-Indo-European patrilateral cross-cousin marriage pattern

Note that in this case $aw^{-1}, aw^{-2}, ego, nepoter^{-1},$ and $nepoter^{-2}$ form a line of 5 matrilineally related males belonging to 5 different generations, a line through which the status and property would move within a matrilineal descent system (see, e.g., Schneider and Gough 1961).

As we see, within such an arrangement ego's MoBr will be SiSo for ego's FaFa. At the meantime ego's SiSo will be MoBr for ego's SoSo. That is, within each couple ($aw^{-1} – aw^{-2}$, and $nepoter^{-1} – nepoter^{-2}$) the senior relative will be MoBr for the junior relative. At the meantime, it seems possible to imagine just one social context when one's FaFa will be denoted with the same kinship term as his MoBr. And this context is nothing else but matrilineal descent. Indeed, within such a context male A would tend to inherit his status and property from his MoBr (B), whereas, as has been shown by Kronenfeld (1991), if A inherits his status and property from B, both of them can be denoted with one kinship term.

In the social context specified above, ego can identify both his MoBr and FaFa (= MoBr's MoBr) as those from whom he inherits his status and property (as Latin data suggest MoBr could still be distinguished from FaFa [= MoBr's MoBr] through the use of a diminutive form [something like *avus* vs. *avunculus*, thus a sort of ‘grandfather’ vs. ‘little grandfather’, i.e. something like a mirror image of English ‘father’ vs. ‘grandfather’]). On the other hand, within the same context an essential characteristic unifying both SiSo and SoSo (= SiSo's SiSo) into one category would be that both of them are ego's heirs.

Note that matrilineal inheritance of status and property is an extremely strong predictor of the presence of matrilineal descent groups (see Tables 8 and 9).
Table 8
Correlation between the Matrilineal Succession to the Office of Local Headman and the presence of Matrilineal Descent Groups (for cultures with hereditary succession)

<table>
<thead>
<tr>
<th>Matrilineal descent groups</th>
<th>Matrilineal succession to the office of local headman</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
</tr>
<tr>
<td>0 = absent</td>
<td>91.4 % (372)</td>
</tr>
<tr>
<td>1 = present</td>
<td>8.2 % (8)</td>
</tr>
<tr>
<td>Total</td>
<td>380</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. p << .0001 (by Fisher's exact test), ϕ = +.76; γ = + 0.983, p << .0001.

Table 9
Correlation between the Matrilineal Inheritance for Real Property and the presence of Matrilineal Descent Groups

<table>
<thead>
<tr>
<th>Matrilineal descent groups</th>
<th>Matrilineal Inheritance for Real Property</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
</tr>
<tr>
<td>0 = absent</td>
<td>94.1 % (491)</td>
</tr>
<tr>
<td>1 = present</td>
<td>4.5 % (4)</td>
</tr>
<tr>
<td>Total</td>
<td>495</td>
</tr>
</tbody>
</table>


Note. The number in parentheses is the n for each cell. p << .0001 (by Fisher's exact test), ϕ = +.806; γ = + 0.994, p << .0001.

The correlations above might appear entirely banal, but in fact they are NOT as banal as one may think. The most relevant point here is that the patrilineal inheritance of status and property predicts the presence of patrilineal descent groups MUCH less strongly (see Tables 10 and 11).
### Table 10

**Correlation between the Patrilineal Succession to the Office of Local Headman and the presence of Patrilineal Descent Groups**  
(for cultures with hereditary succession)

<table>
<thead>
<tr>
<th>Patrilineal descent groups</th>
<th>Patrilineal succession to the office of local headman</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
<td>1 = present</td>
</tr>
<tr>
<td></td>
<td>94.9% (93)</td>
<td>5.1% (5)</td>
</tr>
<tr>
<td></td>
<td>33.0% (134)</td>
<td>67.0% (272)</td>
</tr>
<tr>
<td>Total</td>
<td>227</td>
<td>277</td>
</tr>
</tbody>
</table>


*Note*: The number in parentheses is the *n* for each cell. *p* << .0001 (by Fisher's exact test), $\phi = +.492; \gamma = + 0.948, p << .0001$. SOURCE: Ethnographic Atlas database.

Note that a high value of Gamma in this case is rather misleading, as it reflects the fact that the patrilineal descent is a very strong predictor of patrilineal succession; but on the other hand, the patrilineal succession is a rather WEAK predictor of the presence of patrilineal descent groups with 33 per cent of cultures having patrilineal succession that lack patrilineal descent groups (statistically this is reflected by a relatively low value of Phi-coefficient – in comparison with the rest of correlations above).

### Table 11

**Correlation between the Patrilineal Inheritance for Real Property and the presence of Patrilineal Descent Groups**

<table>
<thead>
<tr>
<th>Patrilineal descent groups</th>
<th>Patrilineal Inheritance for Real Property</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = absent</td>
<td>1 = present</td>
</tr>
<tr>
<td></td>
<td>71.7% (134)</td>
<td>28.3% (53)</td>
</tr>
<tr>
<td></td>
<td>14.8% (65)</td>
<td>85.2% (362)</td>
</tr>
<tr>
<td>Total</td>
<td>197</td>
<td>415</td>
</tr>
</tbody>
</table>


*Note*: The number in parentheses is the *n* for each cell. *p* << .0001 (by Fisher’s exact test), $\phi = +.56; \gamma = + 0.871, p << .0001$.

Hence, the evidence for matrilineal inheritance of status and property among the PIEs may well be regarded as evidence for the presence of matrilineal descent groups among them.
Note that the matrilateral patrilateral cross-cousin model specified above might account for a couple of other PIE kinship terms.

To start with, PIE *g’alow- in addition to the main meaning (‘husband's sister’ has such a reflex as ‘brother's wife’ (e.g., Benveniste 1969, 1973; Gamkrelidze and Ivanov 1995), which might not be a coincidence, as within the model specified above ego's HuSi would tend to be identical with ego's BrWi (see Fig. 2).

![Fig. 2. Identity of Husband's Sister and Brother's Wife within the reconstructed proto-Indo-European patrilateral cross-cousin marriage pattern](image)

Besides this, PIE *dhe:dh- in addition to reflexes FaBr and ‘grandfather’ (in view of the fact that FaFa is denoted as *aw-, *dhe:dh- is likely to have denoted MoFa) has such a reflex as MoSiHu (see, e.g., Benveniste 1973; Gamkrelidze and Ivanov 1995). This might not be a coincidence, as within the model specified above the relationship between the respective kin positions will look as follows (see Fig. 3).

![Fig. 3. Identity of Father’s Brother and Mother’s Sister’s Husband within the reconstructed proto-Indo-European patrilateral cross-cousin marriage pattern](image)
The relationship between MoFa and FaBr would be identical with the one between FaFa and FaBr, i.e., FaBr will be SiSo for MoFa. Hence, within father's matrilineage MoFa's status and property will be inherited by FaBr, that is why they would tend to be denoted by one kinship term (this would be also true for ego's father, but as the kinship terminology we are dealing with is generally bifurcate-collateral, the term for MoFa is merged with the one for FaBr, but not Fa). Note also that within the above specified model ego’s FaBr turns out to be identical with MoSiHu.

Note that the presence of matrilineal descent groups among the PIEs suggested by the analysis of the PIE kinship terminology does not contradict the presence among them of the patrilocal residence suggested by the analysis of the PIE affinal terms (see, e.g., Benveniste 1973; Gamkrelidze and Ivanov 1995), as matrilinearity is perfectly compatible with patrilocal. In fact, in the ethnographic record the number of matrilineal cultures with patrilocal residence exceeds considerably the number of ones with matrilocal residence (see Table 12).

Table 12

<table>
<thead>
<tr>
<th>Matrilineal descent groups</th>
<th>0 = absent</th>
<th>1 = present</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = viri-/parilocal</td>
<td>78.6 % (794)</td>
<td>38.8 % (85)</td>
<td>879</td>
</tr>
<tr>
<td>2 = ambi-/neolocal/no common residence</td>
<td>13.3 % (134)</td>
<td>7.8 % (17)</td>
<td>151</td>
</tr>
<tr>
<td>3 = avunculocal</td>
<td>0.1 % (1)</td>
<td>24.7 % (54)</td>
<td>55</td>
</tr>
<tr>
<td>4 = uxori-/matrilocal</td>
<td>8.0 % (81)</td>
<td>28.8% (63)</td>
<td>144</td>
</tr>
<tr>
<td>Total</td>
<td>1010</td>
<td>219</td>
<td>1229</td>
</tr>
</tbody>
</table>


Of course, this is not the only possible reconstruction of the PIE kinship terminology (and PIE clan-and-family social organization). The aim of the reconstruction described above was to demonstrate that the degree of the Indo-European linguistic reconstruction achieved such a level that it allows suggesting (in coordination with the regularities detected by social anthropologists) rather detailed reconstruction of the clan-and-family organization of the proto-Indo-Europeans. On the other hand, it appears reasonable to note that the most popular reconstruction identifies the PIE kinship terminology as belonging to the Omaha type (see, e.g., Lounsbury 1964; Friedrich 1966; Gates 1971: 73; Barlau 1976; Soseliya 1979; Gamkrelidze and Ivanov 1995: 7.7; Huld and Mallory 1997: 334–335; Kristiansen and Larsson 2005: 5.5; Mallory and Adams 2006: 212; Kristiansen 2009: 122; Friedrich and Pesmen
which, as we could see above, would imply the presence among the proto-Indo-Europeans of patrilineal descent groups.2

Note that such reconstructions can be further verified by using archaeological data.

Indeed, by the moment we have a considerable amount of research devoted to the study of the correlation between certain patterns of kin and family organization and types of dwellings and their patterns that can be detected archaeologically.

For example, of great interest are the correlations between the dwelling floor sizes and postmarital residence patterns detected by Melvin Ember (1973; see also Peregrine and Ember 2002), and further replicated and refined by Divale (1977), Brown (1987), Peregrine (2001), and Porčić (2010). However, the most sophisticated methodology for the identification of the basic features of the kin and family organization of ancient populations on the basis of the archeological data on ancient dwellings’ characteristics and their patterns has been developed in the recent years by Bradley Ensor (2003, 2011, 2012, 2013a, 2013b, 2017).

Note that the data on the ancient dwellings and their patterns that are necessary for the application of Bradley Ensor’s methodology are rather well available in the present-day archaeological record for many archaeological cultures. For example, Stella Souvatzi’s contribution to this special issue (Souvatzi 2017) demonstrates that for Neolithic Greece (6800–3300 BCE) we have archaeological data on the dwelling patterns with that very level of precision that is necessary to apply Bradley Ensor’s methodology to reconstruct kin and family organization of the respective populations.

However, it is not clear at all what was the language of those populations. In fact, according to one of Renfrew’s versions of the ‘Anatolian Hypothesis’ (Refrew 1999) the language spoken by the population of the Neolithic Greece was very close to proto-Indo-European; but there are very strong doubts about the validity of this version of this hypothesis.

Let us recollect at this point that by the moment the two most popular hypotheses of the proto-Indo-European homeland are ‘Steppe Hypothesis’ and ‘Anatolian Hypothesis’.

According to the Steppe Hypothesis the proto-Indo-European homeland is identified with Pontic-Caspian steppes, which more or less corresponds to the areas of the Sredny Stog and Yamna archaeological cultures; according to this hypothesis the split of proto-Indo-European started in the late fifth millennium BCE with the first major migrations out of the Pontic steppes (see, e.g., Gimbutas, 1970, 1977, 1982; Gimbutas, Dexter, and Jones-Bley 1997; Mallory 1989, 1997, 2013; Kortlandt 1990; Anthony 2007, 2013; Anthony and Ringe 2015).

According to the Anatolian Hypothesis (Renfrew 1987, 1999, 2003; Gamkrelidze and Ivanov 1995; Gray and Atkinson 2003; Atkinson and Gray 2006; Gray, Atkinson, and Greenhill 2011; Ryder and Nicholls 2011) the proto-Indo-European homeland was situated in Anatolia, and it
implies that the split of proto-Indo-European began much earlier, around the seventh millennium BCE. According to Renfrew (1999: 266), the split started around 6500 BCE just with the migration of the Neolithic proto-Indo-European agriculturalists to Greece, which would imply precisely that the population of Neolithic Greece (at least in the early Neolithic period) spoke a language that was extremely close to proto-Indo-European (see Fig. 4).

**Fig. 4. Renfrew 1999, 268, Fig. 3 ‘The initial farming dispersal from Anatolia, broadly equivalent to the distribution of Archaic Proto-Indo-European (c. 6500 to 5000 BC) and to the hypothetical Early West Mediterranean Proto-Indo-European (Renfrew’s caption)**

However, there is strong evidence at least against this ‘hard’ version of the Anatolian Hypothesis. For example, there are a number of terms (like axle or wheel) that are securely reconstructed at the proto-Indo-European level (with the Anatolian group being a partial exception). However, the wheeled transport appears to have been only invented around 4000–3500 BCE. This, together with the overwhelming evidence on the importance among the proto-Indo-Europeans of things associated with pastoralist way of life with only occasional agriculture (wool, horses, livestock, dairy foods) seem to point to the Pontic-Caspian steppes as the most likely proto-Indo-European homeland. What is important for us, all such vocabulary is not associated at all with the mode of life of the population of the Neolithic Greece (see Souvatzi 2017).
The Steppe Hypothesis has received recently additional support from genetic research. In 2015 *Nature* published a paper by a team of geneticists titled ‘Massive migration from the steppe was a source for Indo-European languages in Europe’ (Haak et al. 2015). Their analysis of ancient DNA of ancient European populations suggests the presence of massive migrations to Central Europe from the Pontic steppes around 4500 BC. Such a migration correlates to the Steppe Hypothesis of the origins of the Indo-European languages. The authors connect this migration with the bearers of the Yamnaya culture. The study was based on the analysis of 94 genomes of people who lived in the period between 8000 and 3000 BP. The authors conclude: ‘This steppe ancestry persisted in all sampled central Europeans until at least ∼3,000 years ago, and is ubiquitous in present-day Europeans. These results provide support for a steppe origin of at least some of the Indo-European languages of Europe’ (Haak et al. 2015: 207).

Note that these results only destroy a ‘hard’ version of the Anatolian Hypothesis produced by Renfrew (1999, see Fig. 4 above), but not necessarily its softer versions that treat Anatolia as the primary homeland of the proto-Indo-Europeans, and Pontic-Caspian steppes as its secondary homeland (e.g., Cavalli-Sforza 2000; Piazza and Cavalli-Sforza 2006). Indeed, it does not contradict the presence of the wheeled transport terminology in the proto-Indo-European reconstruction. In fact, the main present-day proponents of the Steppe Hypothesis, Anthony and Ringe (2015: 202) concede: ‘Anatolian shares only “thill,” which might be used with a plow or sledge, so Anatolian might have separated before wheels were invented,’ which just implies that ‘post-Anatolian proto-Indo-Europeans’ might have migrated from Anatolia to their secondary homeland well before the invention of the wheel, that is, well before 4000 BCE (leaving Anatolian proto-Indo-Europeans well behind them, in Anatolia). Note that the recent genetic data do not contradict either the ‘soft version’ of the Anatolian Hypothesis. Indeed, the authors of the respective paper note explicitly that ‘the Yamnaya steppe herders… were descended not only from the preceding eastern European hunter-gatherers, but also from a population of Near Eastern ancestry’ (Haak et al. 2015: 207), which turns out to be perfectly compatible with the ‘Second Homeland’ version of the Anatolian Hypothesis.

Note, however, that the abovementioned evidence decreases dramatically the plausibility of the hypothesis that the early Neolithic population of Greece spoke proto-Indo-European.4

On the other hand, some linguistic evidence indicates that at least some part of the Pre-Greek population of Greece may have spoken languages that either belonged to the North Caucasian language family or were directly related to it. Occasional similarities between certain Greek words with no plausible Indo-European etymologies and various Caucasian forms were
already noted by E. J. Furnée in his thorough study of the Pre-Greek substratum (Furnée 1972); later, a more precise series of Caucasian parallels was published by S. L. Nikolayev (1985), who relied on the phonological and lexical reconstruction of Proto-North Caucasian by himself and S. A. Starostin, later published as Nikolayev and Starostin, 1994.

Although not all of the 43 lexical parallels adduced by Nikolayev (1985) are of equal value, there are at least several dozen phonetically and semantically striking look-alikes between Greek and either Proto-North Caucasian or some of the intermediate protolanguages of its daughter branches (such as Proto-Nakh or Proto-Abkhaz-Adyghe) that could suggest Caucasian or 'Para-Caucasian' presence on the Balkans. Some of the most suggestive examples include the Greek word for 'soul' (ψυχή), plausibly comparable with Abkhaz-Adyghe *pa-sʰA 'breath' (cf. especially the reflexion in Ubykh pśá-χʷə ‘to breathe’), Greek ἄχερδος ‘wild pear’ (cf. PNC *qHüːrE ‘pear’), Greek κυπάρισσος ‘cypress’ (cf. PNC *GHab-rišwE ‘gooseberry’, possibly originally generic ‘thorny plant’), and other cultural terms. It remains to be seen whether even more of the so-called ‘Pre-Greek substratum,’ well identified by scholars (see, e.g., Beekes 2014), can be attributed to Caucasian influence, but since possible ‘Caucasoid’ links have been spotted even farther to the west, this looks like quite a promising lead for future research.

Although the North Caucasian hypothesis itself (genetic relationship between Northeast Caucasian, a.k.a. Nakh-Daghestanian, and Northwest Caucasian, a.k.a. Abkhaz-Adyghe) remains somewhat controversial among specialists, largely due to the significant typological discrepancies between the two families, Nikolayev and Starostin have offered a complex historical scenario that reconciles these differences and reconstructs more than 700 common etymologies for Proto-North Caucasian, as well as more than 1000 additional etymologies for Proto-Northeast Caucasian. Lexicostatistical analysis of the basic lexicon that is reconstructable for the Proto-Northwest, Proto-Northeast, and Proto-North Caucasian levels allows to offer a glottochronological dating for the split of the protolanguage somewhere in the time interval of 4,000 BC – 3,500 BC (roughly contemporary with the generally assumed split of Proto-Indo-European), but earlier offshoots of the same linguistic stock may have already existed (and even penetrated Europe) several thousand years prior to that (cf. the Basque-Caucasian hypothesis).

As far as the ‘paleolinguistic’ aspect of correlating reconstructed forms with historical/archeological realities is concerned, it should be noted that lexicon reconstructed for Proto-North or at least Proto-Northeast Caucasian consists of both basic and cultural terms, including a small, but robust core of kinship terms that seems to be less complex (or at least somewhat less historically stable) than the one reconstructed for Proto-Indo-European, but shares occasional curious similarities with it even on the lexical level (e. g.
*nus*ā ‘aughter-in-law’ has been compared by Starostin with Indo-European *snus*- id. as a possible areally diffused cultural term.

Thus, there are some indications that at least some part of the pre-Greek population of Greece might have spoken languages that were rather close to the proto-North Caucasian language. As we have seen, the glotto-chronological analysis suggests that the split of the proto-North Caucasian language took place in the time interval of 4,000 BCE – 3,500 BCE; thus, those fourth millennium BCE inhabitants of Greece who spoke para-North Caucasian languages may have spoken such languages that were rather close to proto-North Caucasian, whose kinship terminology has also been more or less reconstructed.

A substantial contribution to the interdisciplinary analysis of possible linguistic links between pre-Greek populations of Greece and para-North Caucasian language family speakers could be produced by the study of genetic data. Notwithstanding the presence of a number of genetic studies of populations of modern Greece and the Caucasus (see Nasidze et al. 2004; Bosch et al. 2006; Battaglia et al. 2008; King et al. 2008; Zalloua et al. 2008; Balanovskyy et al. 2011; King et al. 2011; Yunusbayev et al. 2011; Roots et al. 2012 and references within them), we do not have yet sufficiently detailed genetic information on those populations whose languages belong to the North Caucasian family. If the future research detects common specific genetic components in the present-day populations of Greece, on the one hand, and in the North Caucasian speaking populations, on the other, dated to the period prior to the disintegration of the North Caucasian linguistic unity – this could serve as additional support for the respective hypothesis. Note that there are grounds to expect that we could detect here a situation that could look as a mirror image of the relationships between the genes and the languages among the Turkic peoples among whom the common specific genetic component (superstratum) has been inherited from the ancestral population together with the language; however, this component only constitutes a minor part of those Turkic speaking populations whose ancestors were assimilated by the Turks (Yunusbayev et al. 2015). Within the hypothesized scenario the general genetic component within the suggested former area of the proliferation of the North Caucasian languages would be a substratum, and the probability of its detection depends on how detailed the genetic description of the respective populations is; it also depends on the contrast range of the genetic background, against which the process of expansion and subsequent contraction of the North Caucasian linguistic area took place.

One may hypothesize that such a genetic substratum may be represented by NRY-chromosome haplogroups widely present in the Nakh-Daghestan and Abkhaz-Adyghe populations, such as various branches of NRY haplogroups G or J and others. The current fast accumulation of the genomic data will allow including in the analysis other genetic systems – up to whole genome sequence.
Hence, the identification of the languages of the inhabitants of the Neolithic Greece still remains problematic.

On the other hand, as has been shown above, there are good indications that the bearers of the Yamnaya culture of the Pontic-Caspi an steppes (starting around 3500 BCE) can well be identified with the speakers of the post-Anatolian proto-Indo-European. However, there appears to be a salient lack of sufficiently high quality settlement plans for this culture that could allow to apply Ensor’s methodology for the reconstruction, with the possible exception of Mikhailovskoe (Lagodovska, Shaposhnikova, and Makarevich 1962). Yet, it is already clear that a sufficient additional work should be done before the abovementioned methodology could be applied to this area.

Still, the International Workshop ‘Murdock and Goody Re-visited: (Pre)history and evolution of Eurasian and African family systems’ that was organized in April 2015 by the Max Planck Institute for Social Anthropology and whose papers have constituted the basis for a special issue of the Cross-Cultural Research (see, e.g., Ensor 2017; Souvatzi 2017) has demonstrated that we are close to having all the necessary ingredients to achieve a new breakthrough in the deep historical reconstruction, whereas Early Neolithic Greece and the Pontic steppes might be identified as ones of the possible springboards for such breakthroughs. Indeed, as has been shown above comparative linguistics and population genetics can potentially provide us with sufficient data to hypothesize about the language spoken by at least some of the populations of such places as Neolithic Greece or the Yamna culture area (including, naturally, the respective kinship terminologies). Those terminologies can be used to reconstruct patterns of kin and family organization practiced by those populations, whereas Bradley Ensor’s methodology presented at the workshop and published in this special issue (Ensor 2017) can be used to find out if those populations actually had those kin and family patterns on the basis of the settlement organization data presented by, say, Stella Souvatzi (2017). This would allow us to identify certain ancient populations about whom we would have not only some substantiated ideas about their basic social organization patterns but also about the language spoken by them, whereas the knowledge of language could provide important clues for the reconstruction of important features of political, economic, cultural and intellectual organization of respective ancient population that can be further verified with archeological, genetic and other data.

Note also, that a fairly detailed reconstruction of the ancient beliefs of human populations is possible by means of quantitative stylistic analysis of modern databases of folklore and mythological motifs (Berezkin 2003, 2005, 2007a, 2007b, 2010, 2015; D’Huy 2012, 2013; Korotayev 2006; Korotayev et al. 2006, 2017). On the other hand, it has been shown that the verification of these reconstructions is possible with the use of genetic data.
Thus, for example, it is possible to reconstruct what the set of beliefs of our ancestors was at the time of their exit from Africa tens of thousands years ago (Berezhkin 2007c). We would add that there are grounds to contend that even certain important features of music and dance styles practiced by ancient populations can be reconstructed through a quantitative analysis of databases or ethnic music or ethnic dance (Lomax 1968). As a result, a synthetic analysis of data from all such sources makes it possible to reconstruct the social and cultural evolution of ancient human populations in an unexpectedly full way; hence, even for a time distance counted in dozens of millennia it appears possible to reconstruct the culture of certain human populations in a very considerable detail including the type of their political organization, the structure of family, beliefs, and even dance or music that those population used.

NOTES

* This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE) in 2019 with support by the Russian Science Foundation (Project No. 18-18-00254).

1 For some detail on the methodology of such a reconstruction see, e.g., Militarev 2010, 2011, 2012.

2 For some other hypotheses about the kinship terminology of the proto-indoeuropeans see, e.g., Kullanda 2002, 2013.

3 For a recent summary of such evidence see, e.g., Anthony 2007, 2013; Anthony and Ringe 2015.

4 Which, however, does not exclude the possibility to test eventually the compatibility of the reconstructed proto-Indo-European kinship terminology with the archaeological data on settlement plans of the populations of Neolithic Greece.

5 Cf., for instance, a list of lexical similarities between Etruscan and Northeast Caucasian, published in Orel and Starostin (1990) (the authors originally proposed to interpret the data as evidence in favor of genetic relationship, but today it seems more reasonable to re-analyze some of those data as traces of old linguistic contacts). Furthermore, cf. various works by J. Bengtson on a possible genetic connection between Basque (Vasconic) and Proto-North Caucasian, evidence for which comes in the form of grammatical evidence and multiple common etymologies in both the basic and cultural sections of the vocabulary, tied together with sound laws (e.g., Bengtson 2008).

6 See Schulze 1997 for a critical review that mentions various problems of the hypothesis, such as difficulties in separating genuine etymological cognates from results of later areal contacts between the various branches of North Caucasian. Nevertheless, for all that concerns historical analysis of comparative Nakh-Daghestanian and Abkhaz-Adyghe data, Nikolayev and Starostin 1994 remains the base reference point for lack of alternatives (Nichols 2003 is an attempt to rework data on Northeast, but not Northwest, Caucasian phonetic correspondences ‘from scratch’ that is much less formalized and detailed).

7 It appears important to note here that the omission of the Anatolian Languages does not appear to affect significantly our reconstruction of the proto-Indo-European kinship terminology.
REFERENCES


Ensor, B. E. 2003. Disproportionate Clan Growth in Crow-Omaha Societies:


Gimbutas, M. 1977. The First Wave of Eurasian Steppe Pastoralists into Copper


