

Comments on Alan Bomhard, “The Origins of Proto-Indo-European: The Caucasian Substrate Hypothesis”



Johanna Nichols
University of California, Berkeley

The paper's main claims are that PIE originated in Central Asia, which accounts for its Eurasiatic properties such as resemblant pronouns (Uralic, IE, Kartvelian, Turkic, Mongolic, Tungusic) and originally agglutinating morphology; then it moved by migration to the western steppe, where profound influence of a North Caucasian language or languages (chiefly West Caucasian) reshaped its sound system, aspects of its morphology, and its lexicon. The work is carefully done, with a large and systematic lexical survey, consideration of archaeological evidence, attention to evidence of contacts and migration, and extensive bibliography. PIE does indeed seem to have a curious typological mix of southwestern and north-central Eurasian traits. I have questions, however, about aspects of the linguistic geography, the Caucasian contacts, and the number and type of lexical resemblances.



Central Asian origin

Bomhard traces the earlier ancestor of Proto-Indo-European (PIE) to a Central Asian homeland to account for its Eurasiatic structural properties,¹ referring to Nichols 1997 and Uhlenbeck 1937 for support. Nichols 1997 proposed that homeland on the evidence of diachronic linguistic geography as it could then be reconstructed, notably a long-standing east-to-west trajectory of language spreads on the steppe and the ability of a Central Asian homeland to explain the southerly parts of Indo-European (in the Transcaucasus and Anatolia) and the geography of the centum-satem split. The paper was first

¹“Eurasiatic” is Greenberg’s term (2000) for a putative macrofamily similar to Nostratic but minus Afroasiatic and Dravidian. I use it here (as I believe Bomhard does) without implying shared descent.



presented in 1990 written up several years before its publication date. Unfortunately, even by the time it came out it was becoming evident that it was grossly inconsistent with very firm archaeological facts and the distribution of PIE terms for wheeled transport and wool technology. (The strongest evidence for the joint implications of wheeled transport and wool vocabulary is laid out in Darden (2001)). I have abandoned that theory and take every opportunity to disavow it. The earliest reconstructable PIE homeland was on the western steppe. There is no reason to posit an earlier origin elsewhere, particularly not to the east; the earlier ancestor of PIE was almost certainly spoken on or near the western steppe, where stockbreeding appeared early and was favored by the ecology and the long familiarity of steppe hunters with the movements and habits of grazing animals.²

The east-to-west language spreads on the Eurasian steppe appear to have begun only in the Bronze Age when mineral deposits in the eastern Urals and the Altai area, together with economic and military growth driven by the mutual dependence of nomads and imperial China, made the eastern steppe a center of expansion and conquest and the cities of the Mediterranean and Near East favored targets of conquest and trade. For the role of China see Barfield (1989). For the linguistic history of the eastern half of Eurasia see Janhunen (1996, 2008, 2012). For the east-west spread of Uralic in this light see Nichols & Rhodes (2018).

Eurasiatic structural properties

Bomhard does not give references that would let me assess the extent to which the archaic morphology reflected in Anatolian points to a Eurasiatic areal context. True, the deeper one goes into Pre-PIE internal reconstruction and morphophonology of inflectional paradigms, the more the morphology approximates the separative, transparent, affixal “agglutinating” type, and the PIE morphology is almost entirely

²Uhlenbeck (1937), Bomhard’s other source for eastern connections of some aspects of PIE grammar, is a perceptive account by an excellent thinker grappling with questions that the knowledge and theoretical tools of the day were unable to handle. (Much the same can be said of Trubetzkoy 1939.)

suffixing. Suffixing "agglutinating" morphology is common in northeastern Eurasia, but both properties are so common worldwide that they are at best only weak evidence of eastern connections.

The resemblances in pronominal roots of IE, Uralic, Kartvelian, Turkic, Mongolic, and Tungusic are striking – at first glance. But the **m- : *t-* pattern of IE and the others is sound-symbolic and similar in typology and favoring context to "mama"-"papa" vocabulary (Nichols 2001); and /m/ proves to be an attractor in small paradigms, more likely to expand than to be lost and susceptible to diffusion under the right sociolinguistic conditions (Nichols 2012a,b, 2018). The combination of sound-symbolic value and attractor status make the Eurasiatic pronominal resemblances a very weak diagnostic of genealogical relatedness (but a tantalizing clue to early Neolithic sociolinguistics and contacts).

Resemblant words

Bomhard presents 164 resemblant lexical items reconstructed for PIE and Proto West Caucasian. The main question is whether these items, with the degrees of semantic and phonological resemblance they exhibit, significantly exceed what could be expected by chance. If they do, we have strong evidence of either contact or common descent; if not, there is no evidence. To assess this I use the criteria of Nichols (2010:305), where table 1 gives the number of matches and degrees of formal and semantic resemblance required to exceed chance in searches through vocabulary sources of various sizes and with various constraints on phonological and semantic distance.

In the terms of that table I judge the 164 items to involve similar consonants without selective parsing, and with semantic ranges from three to five senses (very similar to less similar in the terms of the table). About 120 of them have two consonants, and I will discuss only these. For most of the items the IE and WC sets differ semantically. Some of the differences are what I would call fairly great, e.g. #68 PIE **k^{wh}et^h-* in 'four', which Bomhard reconstructs semantically as 'cut into (equal) parts', and WC (Circassian branch only) **qq^wət^ha* 'smash, break chop'; or #34 IE **g^her-* etc. 'scatter, strew' and WC (Abkhaz branch

only) **yra* ‘speckled, spotted’. By “fairly great” I mean that the two sets of meanings are not entirely unconnected and it is possible to imagine a series of semantic changes that would get from one to the other, but as there is no semantic overlap between the sets of senses there is no obvious reason to assume these changes actually occurred. Some of the differences could be described as semantically close (e.g. with some overlap of likely referents or some partial synonymy), e.g. #29 PIE **d^her-* ‘twist’, ‘turn’ posited as root meaning for cognates ‘twist’, ‘run’, ‘round’, ‘turn’, etc.; and WC **darə* ‘spin, spindle’; for these, semantic change from one to the other seems plausible. I noted two with identical senses: #18 ‘babble, chatter’, etc. and #45 ‘alder’. I estimate a total of about 100 sets with some semantic similarity, divided about equally between semantically close and fairly distant.

I assume the PIE and PWC languages both had some 2000 elementary roots (a typical number, cross-linguistically).³ Especially for PIE, with its extensive history of etymological research, the number of reconstructed roots may be larger, if words attested in only one branch (e.g. #58, Germanic only) or one language (#43, Latin only) are formally reconstructed back to PIE. Bomhard’s survey apparently covered all the PIE and possibly PIE roots, and all the PWC and possibly PWC roots, that could be found in etymological dictionaries, i.e. a search of the full 2000 roots per family. For such a search we need

³Rix 2001 has 1182 PIE roots for verbs. Wodtko et al. 1998 have 207 nominals (nouns and adjectives), a number of which are derived. There are also pronouns, adverbs, and particles. Bomhard consulted all major IE etymological dictionaries and a number of language and branch dictionaries. The total IE roots surveyed must have been at least 2000. I estimate Kuipers 1975 has about 2000 roots for the Circassian branch of WC; the Abkhaz-Abaza branch will add more. Pokorny has ~2000, and that without the additional cognates now available from Anatolian and Tocharian material. Pyysalo 2018 has 730 ultimate recurrent minimal segmentable partials in only items having an Old Anatolian cognate; as single-consonant items these would require twice as many matches to exceed chance. All in all, Bomhard’s careful scouring of many different dictionaries, plus the inherent statistics of protolexicons (whereby any protoform occurs in the protolanguage with 100% probability by definition, while its probability of showing up in any given daughter language is less) and the necessarily broad and approximate nature of proto-meanings, indicate that he had at least the typical ~2000 items to search through in assembling his sets of resemblant forms.



between 121 and 223 resemblants to exceed chance; I assume 170, the mean, rounded off. (As shown in Nichols (2010: table 1, 121) is the number required for very similar senses, 223 for less similar. Again, both are for 2-consonant roots with similar consonants. As a very rough rule of thumb, for a mix of items with fairly similar and fairly distant semantics, about 10% of the items compared need to match.)

In addition there are about 40 one-consonant roots. These require about twice the number of matches to exceed chance, so adding them in will not bring the total appreciably closer to significance.

To summarize, Bomhard's approximately 100 lexical matches falls well short of the needed 170. This is within the range of chance, so the body of material is not good evidence of historical connection.

Most often it seems to be the IE roots that resemble WC roots, but borrowing involves whole words, not bare roots, so it is not clear how borrowing would have occurred. It appears that most of the borrowing was a matter of simplex WC lexemes taken over as PIE roots, which in their attested forms cited by Bomhard mostly have suffixes. A number of the items are verbs, judging from their glosses. The complex polysynthetic verbs of West Caucasian, with their mix of prefixal and suffixal inflection and roots usually consisting of a single consonant, are notoriously hard to segment on the fly and are almost never borrowed into neighboring languages. Even nouns are not frequently borrowed from WC languages, despite the great prestige of Kabardians and Circassians and their customs and society all across the northern Caucasus. In general, the mechanism and sociolinguistics of lexical borrowing in this posited contact episode need a closer look.

Interaction with indigenous Caucasian languages

After the migration to the western steppe Bomhard's model has PIE taking in strong contact effects from interaction with indigenous Caucasian languages, chiefly West Caucasian. However, it is not likely that PIE interacted with early West Caucasian or either of the other indigenous Caucasian stocks. All three indigenous families have their phylogenetic centers of gravity south of the Caucasus. Kartvelian and Nakh-

Daghestanian have arguable homelands to the south (respectively, central Georgia and the Samur delta area on the Caspian coastal plain). The homeland of West Caucasian is unknown. The three branches of the family meet near Sochi, on the Black Sea coast south of the Caucasus, and by center-of-gravity criteria the homeland should be in that vicinity. The Circassian branch (Kabardian and Adyghe, which prior to the Russian conquest of the Caucasus compactly inhabited the northwest Caucasus slopes and nearby lowlands) is not internally deep (less old than Romance or Slavic) and is likely to have spread across the foothills and lowlands into a power vacuum in the wake of Tamerlane's destruction of the Alanic kingdom. On the south slope, Abkhaz is well installed and probably ancient in and near its present range along the western Black Sea coast. There is no reason to think that ancestral West Caucasian occupied the Pontic steppe some 6000 years ago when PIE was still unified. Linguistic diversity in ancient times was greater than now, and we have no evidence of the languages that must have been spoken south of PIE and on the north slope of the Caucasus in ancient times. The most parsimonious assumption is that unknown and unknowable now-extinct languages unrelated to any attested language family occupied the southern periphery of the western steppe. Some of them might have had areal affinities to languages south of the Caucasus and quite likely there were areal and cultural ties to the northern Balkan peninsula.

Summary

Substratal effects are generally thought to involve phonological and grammatical interference but relatively little lexical borrowing (e.g. Thomason & Kaufman 1988, Thomason 2001). Areal contacts often involve similar phonologies, but these can be acquired gradually through diffusion and not only from single intense contact episodes. Bomhard's paper focuses mostly on the resemblant lexical material and resemblant phonological systems, leaving the impression that this was a contact episode but not a classic substratum. My discussion above is concerned with the earlier Central Asian homeland, the diagnostic value of the Eurasiatic resemblances, whether the PIE-PWC resemblances are numerous enough and close

enough to exceed chance, and the general possibility of PIE-PWC or PIE-Caucasian contacts. On all of these counts the evidence is not probative. How PIE acquired its southwestern Eurasian traits when all of western Eurasia north of the Caucasus-Alps-Pyrenees ranges lacks them is a question of major interest, but I doubt that we can answer that question with ordinary lexical and phonological comparison. Indeed, rates of vocabulary loss and change are such that in principle we cannot hope to distinguish genuine sharings from chance at time depths much older than the age of PIE. We need new methods and new questions, but for now we are all in the situation of Uhlenbeck (1937) and Trubetzkoy (1939).

References

Barfield, Thomas

1989 *The Perilous Frontier: Nomadic Empires and China, 221 BC to AD 1757*. London: Blackwell.

Darden, Bill J.

2001 On the question of the Anatolian origin of Indo-Hittite. *Greater Anatolia and The Indo-Hittite Language Family*, 184-228. (Journal of Indo-European Studies Monograph no. 38.) Washington, DC: Institute for the Study of Man.

Greenberg, Joseph H.

2000 *Indo-European and its Closest Relatives*. Stanford: Stanford University Press.

Janhunen, Juha

1996 *Manchuria: An Ethnic History*. Helsinki: Suomalais-Ugrilainen Seura.

2008 Mongolic as an expansive language family. Tokusu Kurebito, ed., *Past and Present Dynamics: The Great Mongolian State*, 127-137. Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa.

2012 The expansion of Tungusic as an ethnic and linguistic process. Andrej L. Malchukov and Lindsay J. Whaley, eds., *Recent Advances in Tungusic Linguistics*, 5-16. Wiesbaden: Harrassowitz.

Kuipers, A. H.

1975 *A Dictionary of Proto-Circassian Roots*. Lisse: Peter de Ridder.

Nichols, Johanna

- 1997 The epicenter of the Indo-European linguistic spread. Roger Blench and Matthew Spriggs, eds., *Archaeology and Language I: Theoretical and methodological orientations*, 122-148. London: Routledge.
- 2001 Why “me” and “thee”? Laurel J. Brinton, ed., *Historical Linguistics 1999*, 253-276. (Current Issues in Linguistic Theory, 215.) Amsterdam: Benjamins.
- 2010 Proof of Dene-Yeniseian relatedness. James Kari and Ben A. Potter, eds., *The Dene-Yeniseian Connection*, 266-278. (Anthropological Papers of the University of Alaska 5 (new series), special issue.) Fairbanks: Alaska Native Language Center.
- 2012a Selection for *m* : *T* pronominals in Eurasia. Lars Johanson and Martine Robbeets, eds., *Copies vs. Cognates in Bound Morphology*, 47-70. Leiden: Brill.
- 2012 The history of an attractor state: Adventitious *m* in Nakh-Daghestanian. Tiina Hyytäinen, Lotta Jalava, Janne Saarikivi, and Erika Sandman, eds., *Per Urales ad Orientem: Iter Polyphonicum Multilingue*, 261-278. (Mémoires de la Société Finno-Ougrienne, 264.) Helsinki: Finno-Ugric Society.
- 2018 Non-linguistic conditions for causativization as a linguistic attractor. *Frontiers in Psychology* 8: 2356.



Nichols, Johanna and Richard A. Rhodes

- 2018 Vectors of language spread at the central steppe periphery: Finno-Ugric as catalyst language. Guus Kroonen and Rune Iversen, eds., *Digging for Words*, 56-68. (British Archaeological Reports (International Series), 2888.) Oxford: BAR Publishing.

Pokorny, Julius

- 1959 *Indogermanisches etymologisches Wörterbuch*. Bern: Francke.

Pyysalo, Jouna

- 2018 Proto-Indo-European Lexicon. Pilot and prospectus: <http://pielexicon.hum.helsinki.fi/>

Rix, Helmut, ed.

- 2001 *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Wiesbaden: Reichert.

Trubetzkoy, N. S.

- 1939 Gedanken über das Indogermanenproblem. *Acta Linguistica* 1.81-89.

Uhlenbeck, C. C.

- 1937 The Indogermanic mother language and mother tribes complex. *American Anthropologist* 39:3.1.395-393.