

Re: Armenian, Sumerian, Burushaski, and Turkic languages

Source: <http://sci.tech--archive.net/Archive/sci.lang/2007-05/msg00806.html>

- *From:* Darkstar <darkstar100@xxxxxxxx>
 - *Date:* 28 May 2007 08:11:38 -0700
-

On May 28, 5:42 am, Nathan Sanders <nsand...@xxxxxxxxxxxxxx> wrote:

In article <1180285900.719794.120...@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>,

Darkstar <darkstar...@xxxxxxxx> wrote:

On May 25, 8:20 pm, Nathan Sanders <nsand...@xxxxxxxxxxxxxx> wrote:

In article
<1180073501.701765.282...@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>,

Trying to derive PIE from English and Spanish would be an
epic
nightmare!

It's good to adjust for the margin of recent phonological deviations
by going to a family proto-state. But you're making it look as if
stand-alone isolates couldn't be compared at all.
If English and Spanish had no family relatives, would we never prove
they are related??

I didn't say it was impossible, only difficult. But the relationship
could not be proven solely by finding a list of superficially similar
words, because just statistically, we expect there to be quite a few
of those just by random chance.

First of all, I'll have to thank you (and Peter T. Daniels) for
indicating the importance of finding regular changes. My only
objection is that you seem to be setting the standards too high – no
regular correspondences have ever been found within some of the well-
established (?) families, like Afro-Asiatic, for instance. If I give
you (say) Dahalo, would you be able to indicate in what way it's

Re: Armenian, Sumerian, Burushaski, and Turkic languages

regularly related to Hebrew? Even in the main Omo–Cushitic group there might be many inconsistencies. And the main IEL's are not entirely unproblematic, either.

But if you say that finding regular laws is absolutely crucial, I'll probably have to agree. The only question is how regular these laws are supposed to be? Can you provide a perfect example of "good laws"? If you use the old IE laws for this matter, it just might turn out that many generally–accepted families simply don't live up to this standard, and by this reasoning will have to discard Afro–Asiatic, Austro–Asiatic, Austronesian, let alone Altaic, or most Amerind families.

What IS the standard that this kind of work is supposed to comply to?

Suppose, I give you some laws, and you or someone else points out a few errors and debunks the whole idea just by saying "these correspondences are not regular enough, period". So which ones are, then?

That is, just on articulatory similarity alone, pure dumb luck causes your method to have 11–33% *accidental* similarity between languages that absolutely nothing at all to do with real genetic relationships.

It's a long story. You make it sound as if the Swadesh lists were totally unsuitable for anything. But that greatly depends on *how* you count the potential cognates. Swadesh can't be blamed if a comparativist is not able to perform phonological analysis. So the 33% you've mentioned one can only get if one is totally impartial or indiscriminate.

As you have been in choosing your "cognates".

1) In practice, there are more places of articulation than "just 3". You're forgetting vowels, semivowels, glottals as well as differences in vibrants, spirants, affricates, etc which all sound different to a discriminate observer.

But you yourself don't make those discriminations, which is why I ignored them in calculating the chance of accidental similarity:

Okay, now I see how you've been counting it. Although I do have a few objections.

You claim erku, alto, and iki as cognates. You have a vowel correspondence set of e~a~i. In other words, you don't care what the vowel is, as long as it exists.

My counter-argumentation for this particular example will be as follows:

- (1) Vowels changes are quite rapid and unpredictable, but I do take them into consideration in these few instances (at least partially).
- (2) There seem to be no other languages in Eurasia which have reconstructed VCCV-structure for "two", and those who do have are located in Central Asia (the above mentioned).
- (3) By performing a careful reconstruction (thanks for pointing out the importance thereof) we have:

(3a) /ichkI/ or /ishkI/ (here, /I/ for a back vowel or shwa) in Turkic on the basis of West Yughur /shigI/, Salar /ishkI/, Yakut /ikki/, Chuvash /ikke/, Uyghur /ikki/ and Tuvan /iyi/ Since there is a law of spirantization in modern Turkic, it could be natural to suppose that it was acting already on the Proto-Tukic level, so the final reconstruction as ***/irkI/* < **/ishkI/* seems plausible.

(3b) /altV/ in Burushaski (the exact ending depends on class) Therefore, we have "erku"-"/altV"-"/ichkI/irkI", which may be quite regular, except for the /t/ in Burushaski, which can possibly be explained as something like **arku* > **arthV* > **altu* conversion in Proto-Burushaski.

Note, that I'm not using any rare or unimaginable sound changes in this example. All changes are explained by reference to the languages I compare, or natural phonological considerations.

Similarly, for "three":

"erek"-"/iski/iske"-"/Ur/Us". The highly raised /U-/ (<u> with the dots) in Turkic indicate that the form in the early proto-Turkic could be something like **/uyr/* or **/eys/*.

But even then, you also have oj, tul, and jilan as cognates, showing that an initial vowel in Armenian doesn't always correspond to an initial vowel in Burushaski and Turkic.

I am not sure about this particular potential cognate. It can also be

Re: Armenian, Sumerian, Burushaski, and Turkic languages

reconstructed as <awdz> (Old Armenian) – "tul/tol" (Burushaski) – /SI la-n/ (Turkic). The Turkic */S/ is known to be a weak palatalized consonant roughly between /s'/, /ch/, /j/ (See for instance O.A. Mudrak "The development of the Proto-Turkic phoneme system"). I don't insist on this particular example, though. It might very well be wrong.

Similarly, you have aun, ha, and ev, which shows that initial vowels in Armenian and Turkic may not necessarily correspond to an initial vowel in Burushaski.

[sigh] If I had been writing more carefully and you had been reading more attentively, you'd know that there seems to be a prothetic /h-/ in Yasin Burushaski which often disappears in other dialects. For this reason, I consider this example to be quite regular.

And you have astl, asii, and sulus as cognates, showing that initial vowels in Armenian and Burushaski may not correspond to an initial vowel in Turkic.

This one's debatable, I agree.

At a closer look, it turns into <astD> (/D/ for unknown coronal, presumably /t/ or /l/) – /asii/ – */Suldur/. And, of course, PIE *ster- (!). There are important similarities in this potential cognate, but the fact that Arm., Greek and Proto-Burushaski tend to have *astr (from "ste-" to "ast-") makes this case more or less irregular.

There is no pattern to your arbitrary groupings. You just find a handful of words that "look the same", without any apparent thought > to systematicity.

No one's analysis is as dumb as just taking vowels, labials, coronals and dorsals in the first syllable.

Look at your correspondences for Armenian, Burushaski, and Turkic.

As stated above, you allow initial vowels of any type to correspond to each other, though this is later contradicted by multiple examples of initial vowels that do not correspond to initial vowels.

You have labials randomly corresponding to other labial: m~m~b, m~b~m, v~b~m, b~b~b, p~b~u.

Ditto for dorsals: k~g~x, k~q~q, k~kh~q.

But bilabials—to-bilabials and dorsals—to-dorsals seem to be regular to me for the purposes of a quick—and-dirty analysis! Especially considering the fact that /q/-to-/k/-to-/x/ and /b/-to-/m/-to-/p/ transition is a natural event in Turkic, which are known to be highly convergent in most other respects! Presumably, there were just 3 or 4 phonemes: */q/ for a weak dorsal, /k/ for a strong one, and */B/ for a weak bilabial. Voiced/unvoiced opposition of these phonemes was irrelevant in Proto-Turkic. (Except for /m/ which does seem to exist in a few cases. As yet, I was unable to differentiate between /B/ and /m/.)

Your coronals are, in theory, split into different groups, but in practice, your groups overlap with each other (and of course, are not consistent within themselves anyway): s~s~j, s~tS~s, j~ts~s, h~t~j, d~O~s, Z~d~t, ts~d~t, l~j~t, l~tr~t, etc.

Coronals do split into different group. And I tried to indicate so in my original post.

Plus, you throw in some random ad hoc correspondences that cross your previous categories, such as h~G~u and j~r~O.

Not random. What you're omitting is that when I do allow for certain inconsistencies, I watch for coincidences in other syllables. For instance:

<hatik> (Armenian) – "GunoG/GonoG" (Burushaski dialects) – /uruG/ (Old Turkic) coincide at three articulation points (!) at the same time (the initial /h-/ is not possible in Turkic).

In addition, you allow for semantic variation (father/man/person, blood/meat, sun/fire/burn, etc.), which exponentially increases the chances of finding accidental similarity: with your lax criteria of (a) unpatterned similarity within (being generous optimistic) 5 similarity categories and (b) semantic variation up to (being generously conservative) 2 meanings deep, any given word has a chance of having an apparent cognate of $1-(1-1/5)^2$, or about 36%.

I allow for semantic variation only when it's semantically appropriate. I do not compare "water" to "fire". Take any good etymological dictionary of any language, there's lots of ALLOWABLE semantic variation.

Re: Armenian, Sumerian, Burushaski, and Turkic languages

In reality, you have only about 3 similarity categories and allow semantic variation up to three meanings deep. This results in a chance of similarity of $1 - (1 - 1/3)^3$, or about 70%!

That's an exaggeration.

And this is just noise due to accidental similarity that has nothing to do with true historical relationships. As long as you allow yourself to use broadly arbitrary correspondences regardless of conditioning environments or consistency within a language (any vowel can correspond to any vowel, any dorsal can correspond to any dorsal, etc.), you're going to have this problem, and you can't escape it until you find *SYSTEMATIC* correspondences.

Agreed. Also, see the initial paragraph.

2) Moreover, I always adjust for lexical stability. Some lexemes are particularly stable (I, not, this, that, thou, mother, water, foot, fauna/flora, numerals),

What do you mean by "stable"?

Sematically stable (in this passage).

Surely not phonetically stable! Sound change is regular, so it affects all words equally. On top of that, the most frequent words of a language are more likely to look irregular in the long run because they are more resistant to morphologically regularizing forces like analogy.

while others can be slangy, compound and semantically vague (fat, good, kill, cloud, rain).

If this means that you intended "stable" to mean "semantically stable", then you're still wrong.

The demonstratives (this/that) easily rise and fall (e.g., in Romance), and some languages don't distinguish between "this" and "that" with single morphemes (e.g., French has *ceci* and *cela*, which are transparently bimorphemic: *ce*=this/that, *ci*=here, and *la*=there).

You might be right about "this/that". Though, the problems you've mentioned do not preclude the reconstruction of *to-/ta in PIE. Nothing is absolutely stable. It's just that the grammatical items exhibit more semantic stability.

Words for animals and plants are unstable. Look at the history and cognates of English deer, hound, sow, worm, cock, daisy, bark, rose, fruit, flower, bush, forest, etc., many of which do not have the same meaning as some of their cognates in related languages and/or have been borrowed.

Let me elaborate. Not ANY fauna/flora terms, but only those that are not similar to others semantically. The name for a particularly zoological or botanical species or genus would usually persist. This principle allows consistent attempts at the PIE homeland reconstruction. Again, nothing is absolutely stable. Just to some considerable extent.

As for "water", you're wrong here too. There are at least two distinct IE cognate sets for "water": PIE wed (English water, etc.) and PIE akw (Latin aqua, etc.).

That is just one well-known exception for "water". Any others in any other languages?

And you've already been shown to be wrong about numbers.

I have been not. Phogland and John Atkinson just gave me two counter-examples. Let's just leave this matter aside. I promise to post some materials (based on Rosenfelder's page) to try to show why I think numerals are generally stable.

2) The phonological stability and the proto-phoneme stability must also be taken into consideration. "Cinco-five" case is so distinct because phonetic monsters like *qwinqwe are highly unstable. But "dos-too" is similar because pure initial coronals rarely dissipate.

Coronals change just as much as any other sounds change. Spanish has [dos], but German has [tsvai], and English has [tu], showing differences in voicing and in affrication. German d- corresponds to English T- (denken~think), but in some English dialects, this has changed to f- (fink). Coronals are particularly vulnerable to

Re: Armenian, Sumerian, Burushaski, and Turkic languages

palatalization (t > tS/S as in nature/nation), and we even find pure coronal to dorsal shifts, such as the Austronesian shift of t to k in Hawaiian.

Different phonemes have different stability. "t/d" in the Anlaut are much more stable than (say) vowels in Auslaut. Is that a news?

If proto-language had many affricates, clusters and vowels, we couldn't expect much correspondence. But if it had almost pure ptk-structure,

"Pure ptk-structure" is pretty rare in the world's language families. Off hand, I can't even think of a major language family that doesn't have at least one member language with doubly-articulated consonants.

There seems to be some kind of misunderstanding of my argumentation. I only said that the a proto-language that has unstable phonemes will yield more divergent offsprings, then a proto-language with a simple and robust ptk-type of structure (consider Japanese as an example of such).

it can survive much longer. For this reason, if you cannot reconstruct ptk-type phonemes, but have some vague correspondence in vowels, spirants, and glotal stops, there's your big red flag.

If you have any "vague" correspondence, you have a red flag. If you can't construct regular, systematic sound changes, you have not identified a genetic grouping.

Okay, I hear you. Or at least I'm trying to.

3) Prothetic phenomena and other types of inclusion are rare.

Hardly! I'd be rather surprised if there was any language in the world that had no epenthesis at all in its history.

You're applying the same faulty logic that you accused me of. "Any language with all of its phonemes in all of its history"! Yes, /h-/ may indeed be on and off sometimes in some languages (I've already mentioned the Yasin dialect of Burushaski). But GENERALLY and STATISTICALLY cases of pure random inclusion of some random phonemes

Re: Armenian, Sumerian, Burushaski, and Turkic languages

are rare. For this reason, explanations of some kind of epenthetic –k– in "duos" with a simultaneous loss of such statistically stable phoneme as the Anlaut /d–/ to explain the origins of "erku" may not work.

4) There is supposed to be some typological similarity in grammar,

Which you haven't even looked at in this case!

There are some questions with the Burushaski four noun class system, but as Peter T. Daniels always says typological (dis)similarities are not crucial ;-). Establishing regular changes is more important – I hope we finally agree on this point.

IMO, only very clear-cut typological dissimilarities should raise objections. When a language is completely dissimilar in many aspects and on many levels. For instance, when trying to compare Khoisan languages to Finno-Ugric.

otherwise you won't be able to reconstruct grammatical morphemes. For instance, the absence of classifiers in English indicates that it would be difficult to relate English to Niger-Kongo.

Nonsense. As I understand it, Tibetan does not have classifiers, but it is definitely related to the other Sino-Tibetan languages (which do have classifiers).

Further, I seem to recall that Old Chinese did not have classifiers, and I don't think we would have much trouble showing that Old Chinese is related to Mandarin.

In many aspects and on many levels, as I've said...

Just think for a second about how overwhelming this can actually be:

we know of sound changes that can cause n to correspond to N, N to g,

g to k, k to ts, ts to s, s to z, z to r, r to l, l to w, w to b, b to p, p to f, f to h, h to ?, ? to t, t to d, d to dZ, dZ to Z, Z to j...

essentially, any consonant in one language could in theory be related

to any other consonant in a different language, given the proper

sequence of sound changes. The classic example of how this

Re: Armenian, Sumerian, Burushaski, and Turkic languages

can happen
in real languages is the pairs of real cognates English "horn"
and
Hindi "singa", or English "five" and Spanish "cinco", which
do not
look like cognates at all.

What you're forgetting is that for infinite changes, you will need
infinite amount of time.

I said nothing about infinity. You only need a few attested sound
changes to go all over the map of the consonant system.

Suppose your proto-sound is *t. Then one daughter language could have
the development t > k (like Hawaiian), while another could have t > ts
(like German).

From these, you could have further developments such as k > h (like
Germanic) and ts > s (like Pipil), resulting in a possible
correspondence set of t~k~h~ts~s after a maximum of two sound changes
per language.

Now, just add in a few more changes, and the diversity of possible
correspondents is staggering. Consider the sound changes h > O
(Romance), O > v or j (Slavic), O > ? (German), s > r (Latin), and r >
l (Samoan). This gives us the 11-member correspondence set
t~k~h~O~v~j~?~ts~s~r~l, all from no more than four sound changes in
any one language.

If this sort of reasoning were right, no proto-language reconstruction
would be possible. You seem to be overdoing it.

Four sound changes in one language's history is a miniscule number in
the time range needed to find a common ancestor for Armenian and
Turkish.

I use only NATURAL, ALLOWABLE or WELL-EXPLAINED phonological changes,
such /b/ being related to /m/ and /p/ already in Proto-Turkic.

Consider Old Turkic /biz/, Yakut /bihi-gi/, Khakas /pes/, Salar /
piser/ but West Yughur /mIs/ for "we". I do not use just ANY changes.

So for a phone to move ten steps away, it
will take ten temporal units,

Re: Armenian, Sumerian, Burushaski, and Turkic languages

False. Sound changes do not happen in regularly-spaced time intervals. You can easily get sweeping changes in just a few centuries (the Great Vowel Shift in English, the period of disintegration in Slavic, etc.).

Aha. Then how many of these "catastrophic", sudden events do we have in the world language history? The logical error here is that you underestimate statistics (just as I do sometimes). In some cases planes do fall down, but generally flying is safe.

Finding correspondences
is only part of the challenge; the next (and most important)
step is
finding the systematic sound changes that create those
correspondences.

Okay. I agree. I do have some laws, they're just not sufficiently elaborated.

That's an understatement!

Thank you for your careful attention to my work. Without such critical and sensible remarks no serious research into any linguistic problems would be possible.

.