Re: Armenian, Sumerian, Burushaski, and Turkic languages

Source: http://sci.tech−archive.net/Archive/sci.lang/2007−05/msg00810.html

- *From*: Nathan Sanders <nsanders@xxxxxxxxxxxx>
- *Date*: Mon, 28 May 2007 13:33:23 −0400

In article <1180365098.014677.199440@xxxxxxxxxxxxxxxxxxxxxxxxxxxxx>,
Darkstar <darkstar100@xxxxxxxx> wrote:

> 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

I didn't set the standards.

> − no
> regular correspondences have ever been found within some of the well−
> established (?) families, like Afro−Asiatic, for instance.

Nonsense. There are certainly plenty of known regular correspondences
within Afro−Asiatic. Do you honestly think that scholars look at
Berber and Arabic and say "oh, they look the same", and leave it at
that?

> If I give
> you (say) Dahalo, would you be able to indicate in what way it's
> regularly related to Hebrew?

If I knew the relevant facts, then yes. If two languages have been
shown to be related, then that means (some subset of) their common
ancestor has been reconstructed and (some of) the sound changes in
their history have been identified.

> Even in the main Omo−Cushitic group there
> might be many inconsistencies. And the main IEL's are not entirely
> unproblemetic, either.

Note that due to sporadic language change (borrowing, analogy, etc.), regularity can be masked. That doesn't mean the regularity isn't there (though it does make it harder to identify).

> 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"?

Read an introductory textbook on historical linguistics. Trask's and Campbell's are both excellent and reasonably priced (~ US$30 each).

> 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?

There is no set standard. You publish your theory, and people either accept it or don't. Further work may or may not be done and new data may or may not come to light that support or detract from your analysis.

> 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?

Well, certainly one example of a correspondence is not enough. A hundred is probably enough. Somewhere in between is the boundary, but the boundary is not a perfect delineation between a proven analysis and an unproven one. It's fuzzy, and different people will have different perceptions of what is "enough" (and the boundary will be different depending on what else is going on in the languages).

If you must have a strict, mathematical answer, you could always construct a computer model of the languages in question and determine how likely it is that the correspondences you found could have arisen by chance. If there is only about a 5% chance or lower that what you have found could have been an accident, then you're usually considered safe. Given how lax your methods are (no systematicity, no regular sound changes, large variance in what counts as "similar"), you're going to be well over 5%.

Another metric you could use is relative to other analyses. You only need to be better than all competing analyses to survive. Since the currently accepted analysis of Armenian is significantly better than anything you have come up with, you are probably barking up the wrong tree by including Armenian.

> 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

This is not a linguistic universal (cf. Polynesian), so you cannot just assume it a priori.

Even in language families with lots of vowel changes, they are not necessarily rapid nor are they truly unpredictable. You simply don't find spontaneous unconditioned changes like a > ü. What you do find are (step−by−step) sound changes constrained by phonetic naturalness.

> 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).

Your suggestion that this is somehow relevant and important means that you believe either that the modern languages have all inherited a VCCV form or that they have a shared innovation that created a VCCV from something else. Are you really prepared to defend either of those claims? Do you have any other evidence whatsoever besides this one word?

And of course, since Armenian "two" is strongly believed not to have come from a VCCV shape to begin with (since it can be derived by regular systematic sound changes from a PIE form that was not vowel−initial), it should not be included in this analysis, until you can come up with an analysis of Armenian's history that is better than what is currently accepted.

> 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.

Do you have any other evidence besides this one word for this analysis? Where else in Turkic do you find S > r?

> (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.

Do you have any other evidence besides this one word for this analysis? Where else in Burushaski do you find k > t and r > l?

> Note, that I'm not using any rare or unimaginable sound changes in this example.

That's not your problem. Your problem is that you come up with sound changes to explain *one word at a time*, with no eye for systematicity. Sound changes are regular, affecting *ALL* words that they could apply to. If you posit an unconditioned change of S to r for Turkic, then *every* instance of S in Turkic should change to r, and you'd need a damn good explanation for why there are still some instances of S floating around.

> > 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!

"Quick and dirty" is not scientific.

> Especially
> considering the fact that /q/−to−/k/−to−/x/ and /b/−to−/m/−to−/p/ transition is a natural event in Turkic,

Turkish has a full contrast between b, p, and m in most positions.

> which are known to be highly
> convergent in most other respects! Presumably,

You presume too much and analyze too little.

> there were just 3 or 4
> phonemes: */q/ for a weak dorsal, /k/ for a strong one, and */B/ for a
> weak bilabial.

What do "weak" and "strong" mean?

> Voiced/unvoiced opposition of these phonemes was
> irrelevant in Proto−Turkic.

What is your source for this claim? Everything I have seen on
Proto−Turkic has posited both p, t, k and b, d, g.

> > 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.

Of course. Semantic drift is a perfectly valid type of language
change. But allowing for semantic drift (which you *have* to allow
for) increases the chances of finding accidental similarities, making
regularity and systematicity all the more important.

> > 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.

It's one of the extreme ends of your method, yes. But given how you are operating, for some words (like "man"), there is about a 70% chance that you find a cognate in any randomly selected language, because you are allowing yourself to look at not just words for man, but also father and human, and you are only considering similarity when the first consonant matches in place of articulation.

And of course, there are vastly more possible semantically related concepts, which ultimately means that, if you look hard enough, you are essentially guaranteed to find a pair of words with man−related meaning that look "similar" in any two languages you look at.

Just consider what's available in English for "man": human, father, leader, chief, adult, person, king, and dude are all only one or two semantic steps away from "man". That gives us the ability to count as apparent cognates any word with a comparable meaning that begins with any of m, h, f, l, tS, Ø, p, k, or d, or any phonetically similar sounds (b, v, r, dZ, S, Z, s, z, g, t)!

> > 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.

Of course, because we have systematic correspondences that help sort through the mess (and an abundance of written records not often found for other languages).

> > Nothing is absolutely stable. It's just that the grammatical items exhibit more semantic stability.

But the words you claim are stable have changed in just a few thousand years in the Romance languages, so why would you suppose without evidence that they remained stable for a longer period of time in a different language grouping?

> > 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?

I have no idea. Without studying the languages in question, I wouldn't want to make claims about whether their word for water was stable or unstable. There's no reason to suppose either way without evidence, because both are possible.

> > 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).

Japanese has affricates.

Aside from the fact that you often get details wrong, I notice that you have the tendency to point to a single example as somehow exemplifying a general rule. Even if Japanese had a "ptk−structure" (which it doesn't) and a stable history (which it doesn't) does not mean that "ptk−structured" languages are generally stable!

> > 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

I'm not talking about "random" sound changes! I'm, as always, talking about *systematic* sound changes.

> 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.

The sound changes deriving Armenian from PIE are not random or arbitrary. They are regular across the language and phonetically natural.

> > 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.

Proto−language reconstruction is possible with *SYSTEMATIC* correspondences. When you ignore systematicity, the wide variety of possible outcomes is what makes the analysis impossible (or more accurately, indistinguishable from random chance).

The point is that you can't just rely on similarity alone, because even for real language change, similarity is not necessarily the result. But when they are dissimilar, they are *systematically*

dissimilar.

> 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.

But you don't use them *systematically*.

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

But from word to word, you do just use "any" change. For one word, you use a m~m~b correspondence, and for another, you use m~b~m. It's completely random which set you use. All that matters to you is that you have some labial corresponding to some other labial, with no concern that you're implying multiple sound changes on the basis of individual words.

> 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?

Who knows? But you can't pretend they don't exist, especially with the time scale and divergent languages you're dealing with.

Nathan

−−
Nathan Sanders
Linguistics Program
Williams College
http://wso.williams.edu/~nsanders/
.