Re: Armenian, Sumerian, Burushaski, and Turkic languages

Source: http://sci.tech−archive.net/Archive/sci.lang/2007−05/msg00772.html

- *From*: Darkstar <darkstar100@xxxxxxxx>

- *Date*: 27 May 2007 10:11:40 −0700

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

> In article <1180073501.701765.282...@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>,
>
> Darkstar <darkstar...@xxxxxxxx> wrote:
>
> > On May 25, 4:46 am, Nathan Sanders <nsand...@xxxxxxxxxxxx> wrote:
>
> > > I mean "Where language 1 has phone X in environments E
> > > and F, language
> > > 2 systematically has phones Y and Z respectively".
>
> > By your tally, you might not be able to prove the relation between
> > English and Spanish.
>
> Of course you can, because it's been done. Hint: you don't start by
> comparing English directly to Spanish. You compare English to the
> other Germanic languages to derive Proto−Germanic, and you compare
> Spanish to the other Romance languages to derive Proto−Romance. Then
> you compare the proto−languages together.
>
> 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 realtives, would we never prove
they are related??

> Have you tried this sort of experiment with your students?

Yes, using the method outlined above.

> Besides, you're forgetting that I'm using Swadesh lists, which are highly stable,

There are well−known problems with blind adherence to the Swadesh lists.

Even if the Swadesh lists were totally viable, with the random tolerance for superficial correspondence you're using, you're predicted to classify up to a third of *any* list as cognate between any two randomly selected languages.

Here's the simplified mathematics: you are only concerned with one or two consonants in the word. To be generous, let's say you're consistently looking at the first two consonants and require them both to be similar between languages. You count consonants as being similar if they have the same major place of articulation (so p, b, m, f, and v all correspond). There are essentially only three major places of articulation (labial, coronal, and dorsal), so each word on the Swadesh list fits into one of 9 (3x3) categories, which means any given word has roughly a 1−in−9 chance of being in the same category of phonetic similarity in two languages (this simplistically assumes the categories are the same size...). Now, since you often only consider one consonant, you can have as large as a 1−in−3 chance of finding similarity that arises randomly.

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 tottaly 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 analisys. So the 33% you've mentioned one can only get if one is tottaly impartial or indescriminate.

Normally, the method yields about 7−15% of background noise (if the comparativist is careless enough). Anything above that can be seen as non−zero correlation.

I've prepared a link with Yoruba−English−Spanish Swadesh lists. There are a few adjustments and refinements to the classical procedure.

http://www.geocities.com/indo_european_geography/Swadesh_lists.htm

Some comment:

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. No one's analysis is as dumb as just taking vowles, labials, coronals and dorsals in the first syllable. (Though, even that is supposed to give 25% of random chance coincidence, not 33%.)

2) Moreover, I always adjust for lexical stability. Some lexemes are particularly stable (I, not, this, that, thou, mother, water, foot, fauna/flora, numerals), while others can be slangy, compound and semantically vague (fat, good, kill, cloud, rain). For instance, "rain" is "falling water" in Turkic. "Tear" is "eyewater" in many Altaic and many SEA languages. But "water" itself is semantically stable, because it's not similar to any other things, it's just a generic name for "something wet". On the contrary, "kill" can be easily substituted by "croak", "ice", "take out", "beat down", "shoot", "murder", etc.
Fauna/flora names are also quite stable. Such words as "moss", "goose", "birch" have very few synonyms (in fact, none) and they are rarely used so they don't get a chance to change, hence Russian "moh", "gus'", "bereza" and Spanish "musgo", "ganso", "abedul".
The high stability of numerals can also be demonstrated (by refering to Rosenfelder's page).
For this reason, I make ordered subgrouping within the main list.

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

3) Prothetic phenomena and other types of inclusion are rare. Therefore, Yoruba "awa" and English "we" might not be cognate, even if they look so. Long dropouts must be accounted for. Therefore, Yoruba "eniayn" and English "all" don't look too similar.

4) There is supposed to be some typological similarity in grammar, 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. Of course, grammatical similarity alone proves nothing. But the absence thereof

may cause difficulties when doing comparison.

Etc.

[...]

> What you have is, roughly, "Where language 1 has phone X, language 2
> randomly has V, W, Y, or Z".

> I say, "Where language 1 has phone X, language 2 randomly has Y1, Y2, or Y3, where X is phonetically related to Y"

> That right there is a problem. Many common sound changes (debuccalization, syncope, apocope, epenthesis, metathesis, rhotacism) or sequences of sound changes (palatalization of k to tS/ts followed by simplification to S/s (as in many IE languages); acquisition of a secondary place of articulation and subsequent loss of the primary place of articulation (as with Polish *l > w); frication of t to T followed by a shift in place from T to f due to acoustic similarity (as in Rotuman); etc.) need not result in obviously related sounds.
>
> In light of the vast range of possibilities that actually occur in real sound changes even over relatively short periods of time, the fact that your method can get so many results just with (place of articulation) similarity should be a big red flag that you're not doing something right.
>
> 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 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. So for a phone to move ten steps away, it will take ten temporal units, which in practice might be as long as 10 thousand years. In a finite amount of time, sounds don't change infinitely.

[...]

> 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.

.