

## **Morphological characters for tracing contact and descent in Nakh-Daghestanian**

Johanna Nichols, UC Berkeley

This paper uses the 18 verb pairs of Nichols et al. 2004 to calculate morphologically based distances in Nakh-Daghestanian along the lines of Nichols 2009a. The 18-pair verb list consists of 18 meanings and their semantic causatives (e.g. 'fear' : 'scare', 'sit down' : 'seat, have sit', '(come to a) boil' : '(bring to a) boil', 'break' (intrans. and trans.), etc. Nichols et al. 2004 used these to typologize the derivational morphology and lexical grammar of languages, based on which type of derivational pairing (e.g. causativization, anticausativization, different light verbs, suppletion, etc.) is used most frequently. The present study uses them to measure distances, in two ways: (1) look at each pair of verbs in each pair of languages, code the derivational types as same vs. different, and calculate the total number of "same" entries for each pair of languages; (2) count not just same vs. different but determine the minimal number of steps required to get from one type to another (e.g. from causativizing to decausativizing is 2 [lose causative, gain decausative]; from causative to ambitransitive is one [lose causative]); calculate the total number of steps for each pair of languages. (2) is analogous to how distances are usually calculated for phylogeny and dialectometry.

This calculation reveals a previously unsuspected cluster of westerly Nakh-Daghestanian languages that have many sharings. Figs. 1 and 2 show raw numbers of sharings (by calculation (1)) among the languages surveyed so far, presented as two sets of verbs which Nichols et al. 2004 found to have typologically different behavior. Here, Fig. 1 shows a strong cluster involving Avar-Andic and Nakh and Fig. 2 shows sparser affinities in the same region. The same cluster is revealed by verb-second tendencies (Nichols 2009b), Type 5 clitics in core chaining (Nichols & Peterson 2010), and transparent gender marking (Nichols 2008). This cluster straddles the deepest genealogical divide in the family -- that between Nakh and Daghestanian -- and so must be due to contact more than descent (though the close affinities within Avar-Andic could also be descent effects). A mystery: the cluster basically reflects the Avar sphere of influence -- except that Lak was in the Avar sphere but is not part of the cluster, while Dargi, more removed from the sphere, does have some affinities to the cluster.

Some languages have many sharings. I counted both overall sharings and out-of-branch sharings, and results were very similar. Chechen and Avar have the most sharings, unsurprisingly, followed by others in the cluster. Xinalug and Lak have the fewest, consistent with their phylogenetic isolation and geographical non-integration; but the degree of non-integration is greater than would have been expected, especially for Lak.

Surveying not the type of derivation but the cognacy or non-cognacy of the verb roots yields a fairly clear picture of the major branches of Nakh-Daghestanian, to be expected as cognacy of hard-to-borrow words should reveal phylogeny well.

Based on all of this I hypothesize that (1) the morphological distances reveal not just any contact but contact episodes intensive enough to produce almost sister-like resemblance in typological features, and (2) the linguistic impact of Avar was stronger to its west than to its east. Also, (3) the sociolinguistic history and prehistory of Lak and the nature of Nakh-Andic interaction are research priorities.

By the time of the conference I will have added more languages (better coverage of Tsezic, Dargi varieties, Lezgian) and finished coding the verb roots and stems for cognacy. Based on preliminary results I expect to find close clustering among Dargi languages and among northern Lezgian language (i.e. clustering among adjacent close sisters as in Andic or Nakh), and changes in mean numbers of sharings.

For the conference, when the survey is more complete, I will also display SplitsTree diagrams and standard phylogenetic calculations.

	Inghush	Chechen	Avar	Godoberi	Karata	Akhvakh	Hinukh	Lak	Dargi	Lezgi	Archi	Xinalug
Inghush	8											
Chechen	6	7										
Avar	6	7	7									
Godoberi	6	7	7	8								
Karata	6	7	7	8	6	6						
Akhvakh	5	6	8	6	6	6						
Hinukh	4	5	5	4	4	6						
Lak	1	2	3	3	3	3	1					
Dargi	3	4	5	5	5	4	4	4				
Lezgi	5	4	3	3	3	3	4	3	3			
Archi	4	5	4	4	4	3	5	2	4	2		
Xinalug	3	2	3	2	2	2	4	0	3	3	3	

Ingush	Chechen	Avar	Godoberi	Karata	Akhvakh	Hinukh	Lak	Dargi	Lezgi	Archi
	9									
	2	2								
	3	3	6							
	1	1	2	4						
	2	2	4	7	4					
	4	4	3	6	5					
	2	2	4	2	1	3				
	2	2	7	5	2	3	3			
	0	0	3	0	1	1	0	3		
	1	1	5	4	2	3	2	2	4	4

Color key:

9-8

7-6

5

Ingush

Chechen

Avar

Godoberi

Karata

Akhvakh

Hinukh

Lak

Dargi

Lezgi

Archi

**Contact:** [johanna@berkeley.edu](mailto:johanna@berkeley.edu) Johanna Nichols, University of California Berkeley,  
Slavic Languages #2979, 6303 Dwinelle, Berkeley CA 94720-2979, USA