Causativization and contact in Nakh-Daghestanian
Author(s): Johanna Nichols
Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society: Special Session on Languages of the Caucasus (2013),
pp. 68-80
Editors: Chundra Cathcart, Shinae Kang, and Clare S. Sandy

Causativization and Contact in Nakh-Daghestanian

JOHANNA NICHOLS
University of California, Berkeley

1 Introduction

In the right sociolinguistic contexts, contact has been found to have predictable typological effects: inter-ethnic languages are less complex while isolated, local, non-interethnic languages are more complex (Szmrecsanyi 2009, Szmrecsanyi & Kortmann 2009, Trudgill 2009, 2011, Ross 1996, Nichols in press). This is because inter-ethnic languages are often simplified in the process of absorbing appreciable numbers of adult second-language learners, while no sociolinguistic force reduces the complexity of non-interethnic languages. Here I report a somewhat different example of such a correlation from the domain of verbal derivation, which has not figured prominently (if at all) in the literature on complexity. The languages surveyed come from the Nakh-Daghestanian (or East Caucasian) family of the eastern Caucasus (Russia, Georgia, Azerbaijan), an excellent natural laboratory for tracking the effects of social context on language structure. This is a very old language family with some 30 named languages and at least 35 actual languages (some of which are called "dialects" but are mutually unintelligible). (1) shows a schematic family tree.

(1) Nakh-Daghestanian subgrouping. * = language with mutually unintelligible dialects.

N-D	Nakh		Chechen*, Ingush, Batsbi
	Daghestanian		
		Avar-Andic	Avar
			Andic: Andi, Karata, Godoberi, Botlikh, Bagwalal, Chamalal, Tindi, Akhvakh*
		Tsezic	Tsez, Khwarshi, Hinuq, Hunzib, Bezhta
		Lak	Lak

Dargwa	Standard Dargi; Akusha, Uraxi, Kubachi, Chiragh, Mehweb, Ic'ari-Sanzhi, Xaidaq'
Lezgian	Lezgi, Tabassaran*, Aghul Rutul, Tsakhur Qrydz, Budukh Archi Udi
Xinalug	Xinalug (possibly a divergent branch of Lezgian)

2 Data and Survey

I survey three general typological properties across all the Nakh-Daghestanian daughter languages for which I could get the relevant data: complexity, transparency, and the number of verb pairs from a fixed list that are derived by overt causativization. Complexity is defined as the number of elements in a system, for a number of different subsystems from across the grammar. The elements and subsystems surveyed here are shown in (2). They are taken from Nichols 2009, with some additional structural variables specific to Nakh-Daghestanian.

(2) Complexity measures. Nakh-Daghestanian measures used here but not in Nichols 2009 are italicized.

Phonology: consonant series, vowels, tones, *phonation types*; syllable complexity
Classification: genders, possessive classes, *noun declension classes*
Inflectional synthesis of the verb: number of inflectional categories (following Bickel & Nichols 2005)
Syntax: alignments, basic word orders
Lexicon: inclusive/exclusive pronouns, *preverb slots, suppletive stems in first and second person pronouns*

The complexity values for the Nakh-Daghestanian languages measured on this scale range from 25 for Lezgi to 51 for Ingush.

Transparency is the extent to which each category has its own discrete marking, in other words the extent to which form mirrors meaning or categories. Kinds of non-transparency include suppletion, allomorphy, and semantically unpredictable categorization. (Transparency in several respects resembles canonical morphology as described for inflectional paradigms by Corbett 2007.) The transparency properties counted here are shown in (3). Transparency is more laborious to determine than complexity, and so far I have surveyed it for only two areas of the grammar.

(3) Transparency measures

Gender: allomorphy of gender markers; allopositionality of gender markers (e.g. prefixal only, prefixal in some verbs but infixal in others, etc.); predictability of gender from semantics

Argument coding: conjunct/disjunct agreement, hierarchical aligment

The gender measures have to do with how transparently the gender of a noun can be determined from its agreement marking and predicted from its semantics. An example of very transparent allomorphy is Avar or Chechen, where all gender agreement markers have a single allomorph, and there are no zero markers and no syncretism. These same languages have no allopositionality of gender markers: the markers on the verb are always root-initial. Less transparent positionality is common among Lezgian languages (e.g. Tsakhur: Dobrushina 1999), where some verbs take prefixal gender agreement and some infixal.[1] An example of transparent gender semantics is Avar, where in the singular all human males take *w*-agreement, human females take *j*-, and all other nouns take *b*-. Another is Tabassaran (Lezgian), where human nouns take *w*- and all others take *r*-. The maximum in semantic transparency of gender is found in the four languages that have lost gender entirely: Lezgi, southern Tabassaran, Aghul, and Udi. A gender system that is partly non-transparent is that of Ingush, shown in (4), where the gender of personal pronouns and human nouns is entirely predictable from person and the sex of the referent, while for other nouns gender classification is arbitrary.

(4) Ingush genders. *v, j, d, b* are the agreement prefixes. Their singular/plural pairings define as many as 8 genders.

	Sing.	*Plural*	*Examples*
1st, 2nd person pronouns	v/j	d	me, you, us
3rd person pronouns (human)	v/j	b	him, her, them
male human nouns	v	b	man, Ahmed (name)
female human nouns	j	b	woman, Easet (name)
some animals, inanimates	b	d	ox, head
some plants, inanimates	b	b	apple, family
inanimates, some animals	j	j	wolf, fence
inanimates, some animals	d	d	dog, house

The argument coding transparency measures have to do with how well one can predict the argument coding from the syntactic categories (or recover the syntactic categories from the coding). Conjunct/disjunct agreement patterns mark person, but mark it non-straightforwardly, wih the same form indicating first

[1] Infixal gender agreement results from entrapment of an agreement prefix when another prefix is added. In Lezgian languages the infixing verbs are generally old bipartite stems that are now largely fused and non-transparent synchronically.

person in questions and second person in statements (or vice versa), so that the relationship between person and marking is not straightforward. Hierarchical alignment marks syntactic relations, but causes argument roles to be obscured or marked indirectly because referential hierarchies (such as animacy or person) determine access to agreement slots.

The verb pairs surveyed are those of Nichols, Peterson, and Barnes 2004, who surveyed a fixed list of verb pairs across 80 languages and typologized languages by their derivational preferences. The verb pairs consist of a verb gloss and its semantic causative, e.g. 'fear' and 'scare, frighten', 'learn' and 'teach', etc. The set of 9 verb pairs surveyed is shown in (5).[2] The possible types of derivational treatment are listed in (6), and the first two are illustrated in (7).

(5) Verb pairs

Plain	*Semantic causative*
laugh	make laugh, amuse
die	kill
sit	seat, have sit, put in sitting position
eat	feed
learn	teach
see	show
be/get angry	anger, make angry
fear	scare, frighten
hide	hide

(6) Kinds of formal derivational treatment of the verb pairs in (5)

Causativization	(semantic causative is overtly derived)
Decausativization	(the plain verb is overtly derived)
Double derived	(both are derived)
Ambitransitive	(neither is derived, as with English *break*)

Ablaut or similar alternation
Change in conjugation class only
Change of light verb
Plain verb is adjective, semantic causative is deadjectival verb
Suppletion

[2] Nichols, Peterson, and Barnes actually surveyed 18 verb pairs, 9 with prototypically animate S/O (e.g. 'fear' : 'scare') and 9 with prototypically inanimate S/O (e.g. 'break', intransitive and transitive). The 9 inanimate ones showed less typological variation and more sensitivity to universals than the others, and since they are all typically of lower text frequency than the animate ones they were less often to be found in dictionaries. These problems obtain for the Nakh-Daghestanian languages as well, so only the first 9, those with animate S/O, are used here.

(7) Examples of causativization (four Nakh-Daghestanian languages, above) and decausativization (two Slavic languages, below). Relevant derivational morphemes bold. Raised "c" = pharyngealization. "lh" = voiceless lateral fricative.

	'fear'	*'scare, frighten'*
Ingush (Nakh)	qier-	qiera-**d.ar**
Hunzib (Tsezic)	hi^{n}ch'a	hi^{n}ch'-**ek'**-a
Avar	h^{c}inq'ize	h^{c}inq'iz-**abize**
Godoberi (Andic)	lhibi	lhib-**al**-i
Macedonian	plaši **se**	plaši
Russian	bojat'-**sja**	pugat'

The languages surveyed are shown in (8). Not all of the Nakh-Daghestanian languages could be surveyed because the grammar survey requires fairly comprehensive descriptions and the survey of verb pairs requires a fairly comprehensive dictionary with an index. Coverage within branches is reasonably good but not optimal. (The Dargwa branch in particular is under-represented.)

(8) Languages surveyed. * = surveyed for verb pairs as well as complexity and transparency (others surveyed for complexity and transparency only).

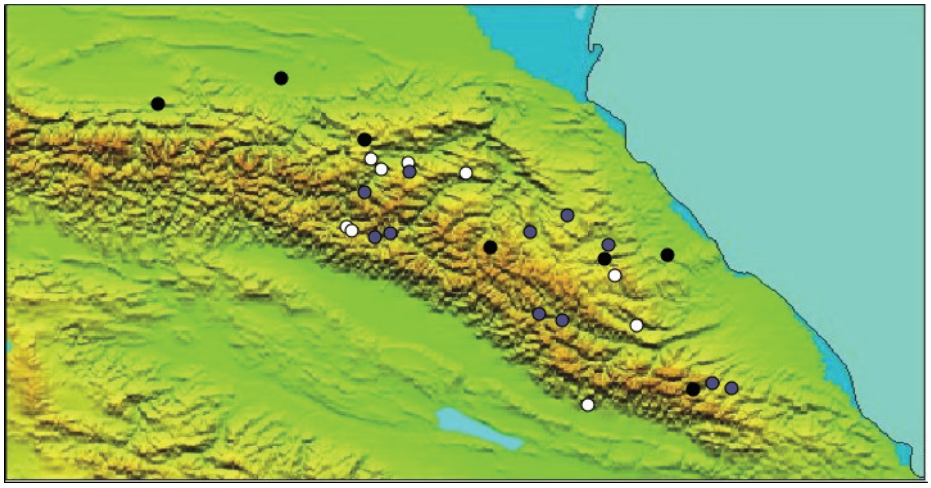
Nakh:	Ingush*
	Standard (lowland) Chechen*
	highland Chechen
Avar:	Standard Avar*
	Antsukh (southern) Avar
Andic:	Northern Akhvakh*
	Karata*
	Godoberi*
	Bagwalal*
Tsezic:	Tsez*
	Khwarshi
	Hinuq*
	Hunzib*
	Bezhta
Lak:	Lak*
Dargwa:	Standard Dargi*
	Kubachi
	Ic'ari
Lezgian:	Lezgi (standard)*
	Tabassaran (northern)*
	Aghul (Bursshag dialect)

Rutul*
Tsakhur*
Qrydz
Budukh*
Archi*
Udi*
Xinalug*

3 Findings

(9)-(11) plot levels of complexity, transparency, and causativization respectively, using the uniform convention that black dots = high values (high complexity, high transparency, high number of verb pairs with causativization), gray = medium, and white = low. The base map shows the eastern half of the Caucasus. Dots are placed at the main town where the language is spoken (for many of the highland languages there is only one town). For languages with large territories (chiefly Ingush, Chechen, Avar, Lezgi) the dot is placed at a historically important or central town (Ongusht for Ingush, Urus-Martan for lowland Chechen, Khunzakh for Avar).

(9) Complexity. White = low complexity, gray = medium, black = high.



For complexity (9), there is no particular distribution to high and medium levels (black and gray dots are scattered throughout the range), but there is a detectable pattern to low complexity (white dots), which forms clusters in two areas. To the west is a cluster of Avar-Andic and Tsezic languages (Avar slightly to the right, the Andic languages Karata, Bagwalal, and Godoberi to its left and at higher altitudes, and the Tsezic languages Tsez and Hinuq in the southern highlands). In geographical terms this means that Avar and nearly all of the Andic and

Tsezic languages along the Andi Koisu river have low complexity. (The remaining Andic language, Akhvakh, is just barely above the cutoff between low and moderate.) To the southeast, three Lezgian languages have low complexity: Aghul to the north, standard Lezgi, and Udi in the southern lowlands.

These two clusters coincide fairly well with the status of (current or past) inter-ethnic language. In the Avar-Andic-Tsezic area we see the results of long-term dominance of the Avar Koisu and Andi Koisu drainages by what is known as the Sarir kingdom (since the early first millennium BCE) and then the Avar Khanate (from its conversion to Islam until the Russian conquest of the eastern Caucasus in 1859). From the capital at Khunzakh (the location of the Avar dot on the map) it dominated the area economically, politically, and linguistically. As was typical throughout the Caucasus, highlanders needed to know lowland languages because the markets and winter pastures were there, while lowlanders did not need to know highland languages. Most of the working-age male highland population was transhumant, spending winters in the lowlands for winter pastures and seasonal work, and spending summers with their herds in highland pastures, and as a result most men were bilingual in their highland language and the lowland language. (See Volkova 1967, Wixman 1980, Nichols 2005.) Consequently, highland villages would occasionally shift to a lowland language, and lowland isoglosses and languages tended to move uphill. Now, the Avar-Andic subgroup is (impressionistically judged) of approximately Germanic-like diversity, and the Tsezic group, whose relatedness to Avar-Andic is widely but not universally accepted, is more divergent.[3] I assume that the stability of the Sarir/Avar dominance meant that there was a long-term spread of language from the Avar lowlands: first Proto-Avar-Andic-Tsezic, then Proto-Avar-Andic, then ancestral Avar spread along similar trajectories, so that the earliest branch to spread and diverge, Tsezic, is now in the highest highlands, the next branch, Andic, is in the lower highlands, and Avar has dominates the foothills, lowlands, and main river canyons. Human habitation of the Daghestanian highlands goes back millennia earlier, so each uphill spread replaced previously present languages by language shift. Thus at all times in the process the language spreading uphill was an inter-ethnic language much as Avar has been in historical times. The low complexity of most languages in the Avar sphere is consistent with their having been inter-ethnic languages. In the case of the Andic and eastern Tsezic languages the low complexity survives although the languages have not had inter-ethnic status for centuries or even millennia.

To the southeast the picture is probably similar. Lezgi and Aghul are very closely related.[4] Lezgi is a large language with a large speech community and large range, and is an inter-ethnic language in market towns in the nearby up-

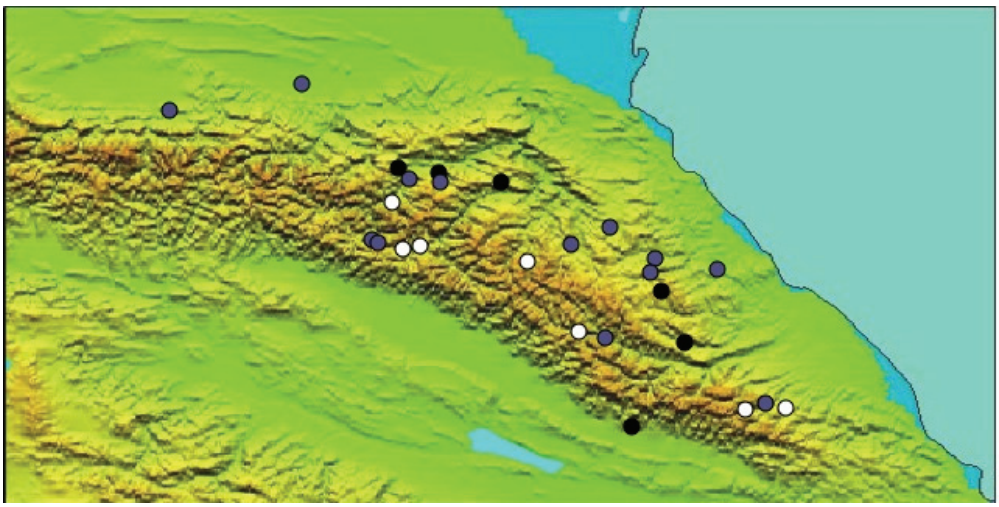
[3] Korjakov 2006:21, 28 dates Avar-Andic to about 3500 BCE and Avar-Andic-Tsezic to slightly earlier, based on glottochronological counts that I have not reviewed.
[4] Korjakov 2006:21 dates this branch at about 2500 years old, though see again note 3.

lands. Less is known about the ethnohistory of this area, but there is a long history of states and kingdoms dominating the Caspian coastal plain (which widens out in the Lezgi lowlands). Lezgi extends along the lower and middle Samur (the major river in the area) and its tributaries, and Aghul is adjacent and just above it on two major tributaries. This is a likely result of long-term spreading from the Samur lowlands, and the history of statehood means that there were important economic centers on the plain from which any major spread along these rivers must have emanated. Another close sister of Lezgi and Aghul, Tabassaran, is nearby as the crow flies but centered on a different river system and therefore probably not a result of the same spread. Its complexity is high.

Udi, the southernmost language and the only one centered on the south slope of the Caucasus range, is best known to linguists for its endoclisis (Harris 2002). It is a small language, spoken in two towns in Azerbaijan and a recently formed outpost in Georgia. Though an endangered enclave language now, Udi in the mid first millennium CE was an important language of the south Caucasian lowlands, known to philology as Alwan or Caucasian Albanian. It had a script created for it by Byzantine Christian missionaries, was an inscriptional language, and has a gospel translation in a recently discovered and published palimpsest (Gippert et al. 2009). Its low complexity is consistent with this history and, if that is its explanation, has lasted over a millennium.

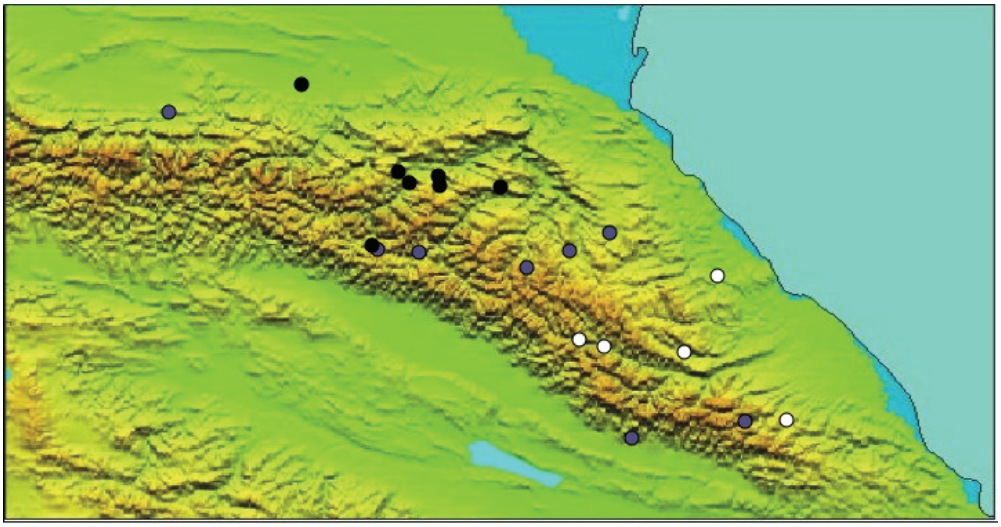
Transparency (10) gives a comparable picture. Low transparency (white dots) is found only in the highlands, and though the topography on the base map used here does not show this clearly most of the low-transparency languages are at the highest inhabited levels on their watercourses. These are languages that are sociolinguistically and geographically isolated and no known history of inter-ethnic use. High-transparency languages (black dots) again cluster in the Avar sphere and the Lezgian area, plus Udi.

(10) Transparency. White = low, gray = medium, black = high.



Causativization of verb pairs (11) shows a different distribution. Black dots are languages in which high numbers of the verb pairs are causativized. They form a large cluster in the Avar sphere, where Avar and some of the Andic languages have some of the world's highest proportions of causativized verbs (Creissels 2009), and they also include adjacent lowland Chechen. White dots have low numbers of causativized verbs. They include most of the Lezgian languages, all of which make extensive usage of light verb constructions and form many of their plain-causative pairs by using two different light verbs. Light verb constructions are frequent in the area comprising the southeast Caucasus, northern Azerbaijan, and nearby (Stilo 2009). Light verb constructions are a frequent typological correlate of the kind of lexicon in which simplex verbs are a closed class, and this too dominates in the same area. Most and perhaps all of the Nakh-Daghestanian languages have a closed class of simplex verbs and derive and borrow new verbs chiefly by forming light verb constructions. Therefore it is likely that the ancestral Nakh-Daghestanian type had a closed verb class and used an appreciable proportion of light verb constructions in its causative verb pairs, so that the preference for causativization in the Avar area is innovative.

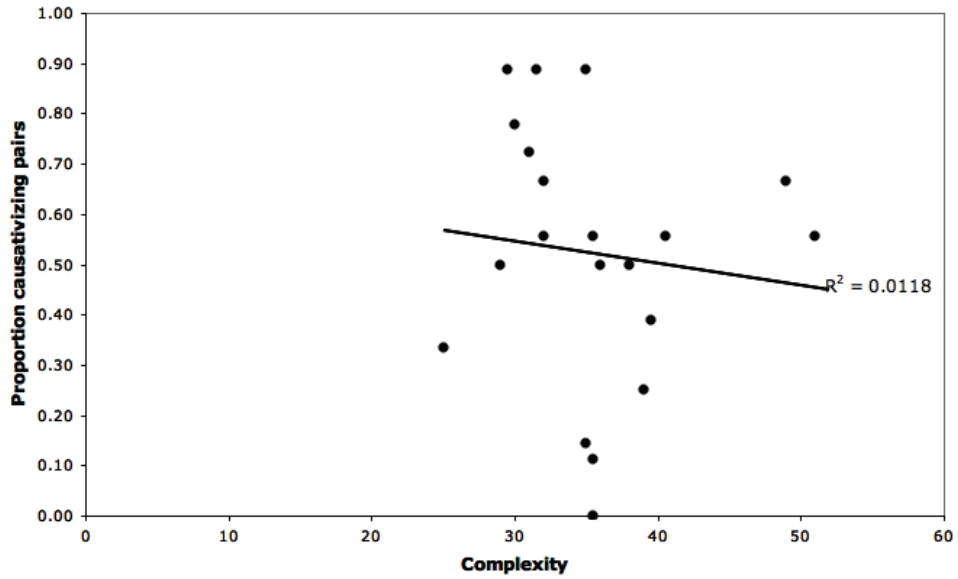
(11) Proportion of the 9 verb pairs that are causativized



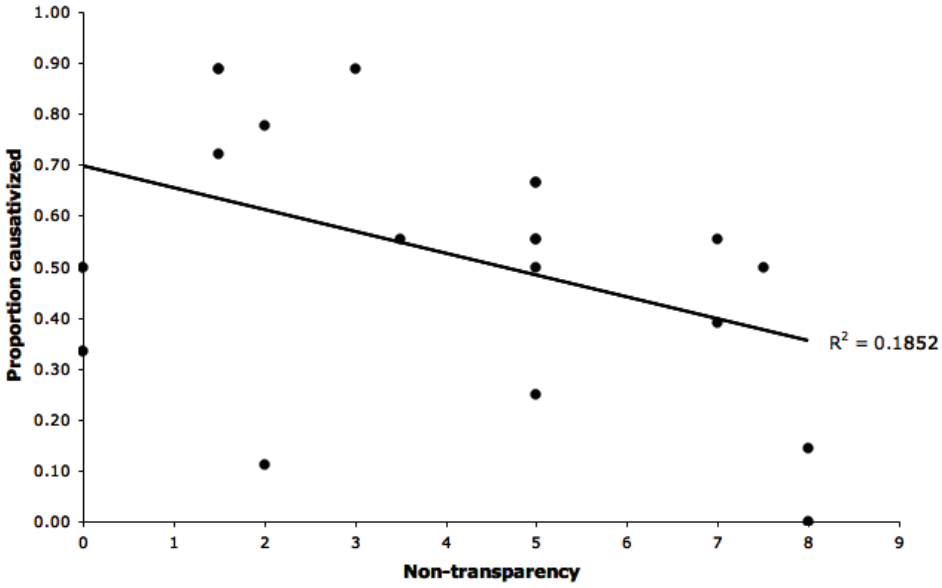
(12)-(13) show how complexity and transparency correlate with causativization, not in geographical clusters as just discussed but language by language. There is essentially no correlation with complexity; the trendline is nearly level. There is an appreciable correlation with transparency (measured as non-transparency in (12), so the negative correlation of causativization with non-transparency is a positive correlation of causativization with transparency). But even this correlation is not particularly strong, which suggests that it is not a purely typological correlation but rather is due to something in the historical

contingencies of the two clusters of languages.

(12) Complexitivity and proportion of pairs causativized. (N = 20)



(13) Non-transparency and proportion of pairs causativized. (N = 20)



The causativizing type is itself an instance of high transparency. The usual analysis of pairs like 'fear' and 'scare' in syntax and semantics is that 'scare' consists of 'fear' plus causation. Assuming this reflects linguistic reality, then a morphological structure {fear-CAUS} is maximally transparent, and a lexicon that

uses this structure widely is not only transparent but also non-complex in that the majority of verbs follow the same pattern. I suggest that this type has increased over time in the Avar sphere as repeated contact among the many small languages there, and contact of all of them with Avar, has made models of the transparent structure available and has favored borrowing and calquing of those models. Note that in the Avar area the smaller communities are politically autonomous and their languages are well retained; Avar is a lingua franca and there has been some shift to Avar but no whole-scale shift. An Avar variety was the language of command in the Avar army, but it was never imposed as a state or official language. The range of Avar as lingua franca extends well beyond its range as first language.

The sociolinguistics is different in the southeast, where causativization is not frequent. There, it is not that Lezgi serves as lingua franca among many small speech communities whose own languages remain autonomous; rather, if there were such languages, they have shifted to Lezgi, and the range of Lezgi as inter-ethnic language (one cannot really call it a lingua franca) does not extend far beyond its range as first language. The size of Lezgi is then due to an ordinary language spread, while the Avar-Andic-Tsezic region is an area of stable, long-term, and complex multilingualism. Both the Avar and the Lezgi situations have favored reduction of complexity and increase of transparency, but only the Avar one has fostered the long-term lexical influence that has favored spread of the most transparent model of verbal derivation, perhaps one word at a time (calque, loan translation, loan), gradually building up consistency across the lexicon.

There is some circumstantial evidence in favor of this analysis. There are three other areas in the world where strongly causativizing languages cluster: the eastern Eurasian steppe and nearby (Turkic, Mongolian, Tungusic, Tibetan); the Austronesian languages and some of their neighbors; and western North America (Nichols, Peterson, Barnes 2004, Fortescue 1998). These are all places where complex contact patterns involving many languages, probably including multilingualism and/or back-and-forth local language shifts, were the rule for long periods. If I am right about the Avar area, expansions of transparent patterns across the grammar and lexicon could have occurred in these places as well.

Thus the extralinguistic situation can plant seeds of change that interact with typological pressures to eventually produce an unexpected cluster of languages that are not true to their family's type.[5]

[5] Some of the work reported here was supported by NSF BCS 9222294, 9606448, and 0966675. Some of the research was done in the Max Planck Institute for Evolutionary Anthropology, Leipzig.

References

Bickel, Balthasar and Nichols, Johanna. 2005. Inflectional synthesis of the verb. In Martin Haspelmath, Matthew Dryer, Bernard Comrie, and David Gil, eds., *World Atlas of Language Structures*, 94-97. Oxford: Oxford University Press.

Corbett, Greville. 2007. Canonical typology, suppletion, and possible words. *Language* 83:1.8-42.

Creissels, Denis. 2010. P-lability and radical P-alignment. Presented at 19th International Symposium on Theoretical and Applied Linguistis, Workshop on Typology of Labile Verbs: Focus on Diachrony. Aristotle University of Thessaloniki, April 3-5, 2009.
Online: http://www.deniscreissels.fr/public/Creissels-P-lability.pdf
(accessed October 2011)

Dobrushina, Nina. 1999. Glagol. In A. E. Kibrik, ed., *Èlementy caxurskogo jazyka v tipologicheskom osveschenii*, 58-85. Moscow: MGU.

Fortescue, Michael. 1998. *Language Relations across Bering Strait*. London and New York: Cassell.

Gippert, Jost, Wolfgang Schulze, Zaza Aleksidze, and Jean-Pierre Mahé. 2009. *The Caucasian Albanian Palimpsests of Mount Sinai*. Turnhout, Belgium: Brepols.

Harris, Alice C. 2002. *Endoclisis and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

Korjakov, Ju. B. 2006. *Atlas kavkazskix jazykov*. Moscow: Piligrim.

Nichols, Johanna. In press. The vertical archipelago: Adding the third dimension to linguistic geography. In Peter Auer, Martin Hilpert, Anja Stukenbrock, Benedikt Szmrecsanyi, eds., *Space in Language and Linguistics*, Berlin: Mouton de Gruyter.

Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 110-125. Oxford: Oxford University Press.

Nichols, Johanna. 2005. The origin of the Chechen and Ingush: A study in alpine linguistic and ethnic geography. *Anthropological Linguistics* 46.129-155.

Nichols, Johanna, David A. Peterson, and Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8:2.149-211.

Ross, Malcolm D. 1996. Contact-induced change and the comparative method: Cases from Papua New Guinea. In Mark Durie and Malcolm D. Ross, eds., *The Comparative Method Reviewed*, New York: Oxford University Press.

Stilo, Don. 2009. An introduction to the Araxes-Iran linguistic area. MS, Max Planck Institute for Evolutionary Anthropology, Leipzig.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21:3.319-354.

Szmrecsanyi, Benedikt and Kortmann, Bernd. 2009. Between simplification and complexification: Non-standard varieties of English around the world. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 65-79. Oxford: Oxford University Press.

Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil, and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*, 98-109. Oxford: Oxford University Press.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.

Volkova, N. G. 1967. Voprosy dvujazychija na Severnom Kavkaze. *Sovetskaja ètnografija* 1967:1.27-40.

Wixman, Ronald. 1980. *Language Aspects of Ethnic Patterns and Processes in the North Caucasus*. Chicago: University of Chicago.

Johanna Nichols
University of California, Berkeley
Department of Slavic Languages and Literatures, m.c. 2979
6303 Dwinelle Hall
Berkeley, CA 94720-2979, USA

johanna@berkeley.edu