

Word-order variation in a contact setting: A corpus-based investigation of Russian spoken in Daghestan

CHIARA NACCARATO* AND ANASTASIA PANOVA

Linguistic Convergence Laboratory, National Research University Higher School of Economics

NATALIA STOYNNOVA

V. V. Vinogradov Russian Language Institute of the Russian Academy of Sciences / Linguistic Convergence Laboratory, National Research University Higher School of Economics

ABSTRACT

This paper deals with word-order variation in a situation of language contact. We present a corpus-based investigation of word order in the variety of Russian spoken in Daghestan, focusing specifically on noun phrases with a genitive modifier. In Daghestanian Russian, the nonstandard word order GEN+N (prepositive or left genitive) often occurs. At first glance, this phenomenon might be easily explained in terms of syntactic calquing from the speakers' left-branching L1s. However, the order GEN+N does not occur with the same frequency in all types of genitive noun phrases but is affected by several lexicosemantic and formal features of both the head and the genitive modifier. Therefore, we are not dealing with simple pattern borrowing. Rather, L1 influence strengthens certain universal tendencies that are not motivated by contact. The comparison with monolinguals' Russian, in which prepositive genitives sporadically occur too, supports this hypothesis.

LANGUAGE CONTACT AND WORD-ORDER VARIATION

In the literature on language variation in contact settings, there are more and more frequent discussions on whether phenomena that could be easily explained as the result of language contact are indeed motivated by other factors (cf., among others, Poplack & Levey, 2010; Thomason, 2009). Within this discussion, we might distinguish between two types of approaches. Approaches of the first type are empirically oriented, include huge amounts of corpus data, and use quantitative methods to analyze them. These approaches allow the researcher to account for both contact and noncontact factors without any bias. They mainly arise within the framework of variationist sociolinguistics (cf., Labov, 1972). Some examples of this type of approach to the study of contact-induced features

*Corresponding author. E-mail: ch1naccarato@gmail.com

in morphosyntax are the papers by Mougeon, Nadasdi, and Rehner (2005) and Poplack, Zentz, and Dion (2012); see also Poplack (2021) for a general overview.

Approaches of the second type are typologically oriented, and are used, for instance, to reveal the range of shared features motivated by language contact in large linguistic areas (e.g., Aikhenvald & Dixon, 2007; Hickey, 2017; Matras, McMahon, & Nigel, 2006; Muysken, 2008; Nichols, 1992). Such approaches distinguish cross-linguistically common phenomena, for which language-internal explanations can be suggested, from phenomena that are rarer or nontypical for a given language family. For the latter, it is more plausible to assume a pure contact motivation (e.g., Dahl, 2001:1457; Heine & Nomachi, 2013:77). On the other hand, established typological patterns supported by more general functional (cognitive) motivations are reported to be borrowed and diffused more easily (e.g., Aikhenvald, 2007:31) and are not likely to be replaced by more rare patterns in a contact situation (e.g., Campbell, 1997:51). Bickel (2017) discussed both sides of the interplay between cross-linguistic universals and areal (contact-influenced) patterns in terms of “functional and event-based triggers” of language change.

The two approaches are rarely combined. One exception is represented by the framework of dialect (or sociolinguistic) typology (cf., for example, Kortmann, 2004; Röthlisberger & Szmrecsanyi, 2019), which, however, concentrates mostly on English varieties. This paper contributes to the discussion on contact-induced versus noncontact-induced linguistic variation by relying on both approaches for the study of a specific topic, that is, word-order variation in genitive noun phrases in the contact-influenced variety of Russian spoken in Daghestan.

Although syntactic changes usually start at an intense stage of contact, word order is one of the linguistic features said to be more easily borrowed (Thomason, 2001:70-71; Thomason & Kaufman, 1988:74-76). In some cases, language contact does not lead to the emergence of a new word-order pattern but to the reinforcement of a pattern that already exists in the target language as a marginal one (Heine, 2008:55). Nevertheless, most previous investigations on contact-induced word-order variation in the world’s languages focus on variation at the sentence level (as in the examples listed in Heine [2008:34-35] and Thomason & Kaufman [1988:55]). Those dealing with variation within the noun phrase are definitely less numerous, though some consider word-order variation in noun phrases with a genitive modifier, including in contact varieties of Slavic languages (see Sussex [1993] on émigré Polish, and Leisiö [2000, 2001] on two varieties of Finnish Russian; see also some considerations in Suleymanov [2016] on Baku Russian).

As for the variety of Russian spoken in Daghestan, one of the most prominent syntactic features is the contact-induced shift to left branching in some constructions, including noun phrases with a genitive modifier (Daniel, Dobrushina, & Knyazev, 2010:85-86). Whereas in monolinguals’ Russian the neutral and most frequent word order in noun phrases is N+GEN

(“postpositive” or “right” genitive), in Daghestanian Russian the opposite order GEN+N (“prepositive” or “left” genitive) often occurs, as in (1) and (2).

- (1) *Eščë mat' moej svekrovi*
 also **mother my.F.GEN mother-in-law.GEN**
byla krepostnaja
 COP.PST.SG.F serf.F
 ‘And also **my mother-in-law’s mother** was a serf.’ [RNC, Memories]¹

- (2) *Ona vsju žizn' ne ljubila*
 3SG.F all.F.ACC life.ACC NEG love.PST.SG.F
deduški sestru
grandfather.GEN sister.ACC
 ‘In her entire life she never loved (**her**) **grandfather’s sister**.’ [DagRus, L1 Avar]

The frequent occurrence of cases like (2) in Daghestanian Russian could be explained in terms of syntactic calquing from the speakers’ first languages (L1s), all featuring a neutral GEN+N order in noun phrases (Testelec, 1998:274), as shown in (3), from the Andalal dialect of Avar.

- (3) *λ'ε de-χε mahandi-l q'alam!*
 give.IMP 1SG-LOC **Muhammad-GEN pencil**
 ‘Give me **Muhammad’s pencil!**’ [From (Testelec, 1998:274)]

However, the GEN+N pattern does not occur equally in all types of noun phrases in Daghestanian Russian and correlates with several lexicosemantic and formal features of both the head and the genitive modifier. This cannot be straightforwardly explained as the result of L1 influence.

Moreover, there are deviations from the basic word order in monolinguals’ Russian too, which makes the picture far less trivial than it could look at first sight. Examples of the GEN+N pattern in monolinguals’ Russian are often associated with a specific information structure and have a marked intonation pattern, as in (4): the prepositive genitive modifier takes the phrasal accent (Zemskaja, 1987:151).²

- (4) *sejčas-to imeni Nikolaj Petroviča*
 now-PTC **name.GEN Nikolai Petrovich.GEN**
Osipova orkestr
Osipov.GEN orchestra
 ‘Now there is **the orchestra named after Nikolai Petrovich Osipov**. (Before there was a similar orchestra named after Andreev.)’ [RNC, Memories]

Therefore, we cannot assume that the occurrence of nonstandard word order in Daghestanian Russian is purely a matter of calquing from the speakers’ L1s, as it is often taken for granted in studies on similar phenomena in other contact situations. Our quantitative corpus-based investigation of noun phrases with a genitive

modifier in Daghestanian Russian as compared to monolinguals' Russian points to several factors inducing the occurrence of nonstandard word order, contact being only one of the components involved.

In this paper, we first illustrate our corpus-based analysis of noun phrases with a genitive modifier in the speech of bilinguals from Daghestan and discuss the impact of different parameters based on statistical analyses. Then we compare bilinguals' and monolinguals' speech, based on data from Russian monolinguals' spoken texts. In the final sections we summarize and discuss our findings also in light of some cross-linguistic considerations and draw our conclusions.

NOUN PHRASES WITH A GENITIVE MODIFIER IN
DAGHESTANIAN RUSSIAN: DATA SELECTION AND
ANNOTATION

Following Daniel et al. (2010), we use the term “Daghestanian Russian” to refer to the variety of Russian spoken in Daghestan. Although we are aware of the fact that Daghestanian Russian is actually a set of varieties which diverge from one another depending on the speaker's place of origin, L1, year of birth, etc., we will employ this term to refer to the set of these varieties as a whole. For this specific study, it seems particularly convenient to adopt this notion because the phenomenon under investigation is apparently found in all varieties of Russian spoken in Daghestan.

In the Republic of Daghestan, over forty indigenous languages are spoken. Most of them belong to the Nakh-Daghestanian (East-Caucasian) family, but there are also three Turkic languages (Kumyk, Nogai, and Azerbaijani), and one Iranian language (Tat). Over the last few decades Russian has become the interethnic lingua franca of the region (Dobrushina & Kultepina, 2021). Nowadays it is either a second language (L2)—it is certainly so for most inhabitants of the mountain villages—or the urban vernacular spoken as an L1 by those younger people who were born in towns (Daniel & Dobrushina, 2010; Daniel et al., 2010). As a result of the intense contact with local languages, Daghestanian Russian has developed a number of features at different levels: phonetics, morphology, syntax, and the lexicon (Daniel & Dobrushina, 2010, 2013; Daniel et al., 2010).

The data used in this study were taken from the Corpus of Russian spoken in Daghestan (Dobrushina, Daniel, von Waldenfels, Maisak, & Panova, 2018), which contains a collection of sociolinguistic interviews with inhabitants of rural areas of highland Daghestan for whom Russian is an L2. Most of them learned Russian at school, they are currently fluent in Russian, but use their ethnic language to a greater extent. The interviews were collected in the 2000s in fieldwork trips to nineteen villages in which thirteen languages are spoken, see Figure 1.

At the moment of data extraction, the corpus contained about 186,000 words. First, we extracted all examples that had been tagged by the morphological parser embedded in the corpus (TreeTagger) as nouns in the genitive (2,412 occurrences). Then we selected those examples in which the genitive was part of

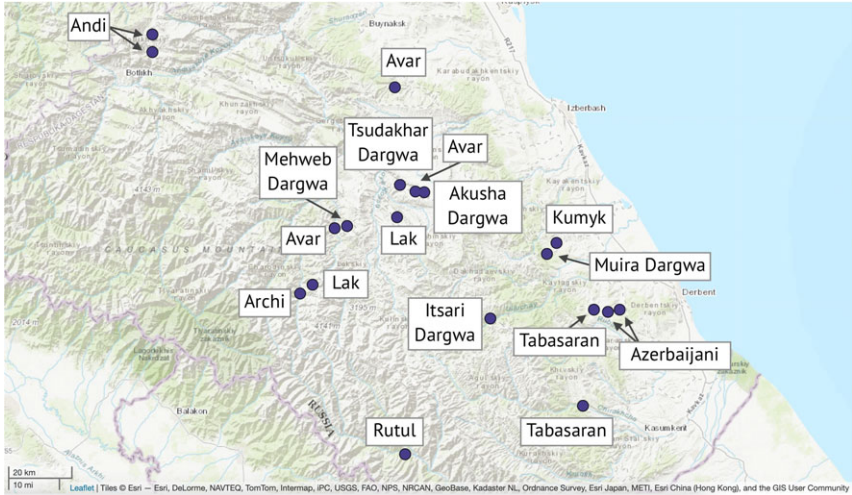


FIGURE 1. Speakers' L1s in the DagRus corpus: geographical distribution.³

a noun phrase (798 occurrences). We further excluded noun phrases whose heads show syntactic and semantic features that make them closer to other constituent types and that are realized with different syntactic structures in the speakers' L1s. Such occurrences include:

- nouns semantically similar to numerals (e.g., *kuča deneg.GEN* 'a lot of money'), and other "quantitative" constructions (e.g., *kusok zemli.GEN* 'a piece of earth');⁴
- constructions with a prepositional function (e.g., *v storonu morja.GEN* 'seaward, lit. to the side of the sea');
- examples featuring a predicative noun as their head (e.g., *okončanie instituta.GEN* 'graduation, lit. the completion of the institute').

We also excluded conventionalized expressions (e.g., *kandidat nauk.GEN* 'Candidate of Sciences [PhD]') because, due to their "frozen" status, they are less sensitive to word-order variation.⁵ We were left with 482 examples, 103 of which exhibited a left genitive (21.4%).

If the occurrence of nonstandard word order in Daghestanian Russian was purely a matter of syntactic calquing from the speakers' L1s, we would expect left genitives to be uniformly represented in our data, because the GEN+N order is the unmarked option in the languages of Daghestan. However, our data show asymmetrical distribution in different contexts, and factors other than contact are needed to account for this asymmetry.

Thus, we analyzed our data according to a number of parameters, each of which is discussed in the rest of this section.

Sociolinguistic parameters

Each example in our corpus was tagged according to the following features: the speakers' L1 and place of birth, year of birth, gender, and level of education.

The data come from forty speakers, twenty-six men and fourteen women, born in nineteen different villages and speaking thirteen different L1s. L1s can be clustered into five genetic groups: (1) Avar-Andic (Avar, Andi); (2) Dargwa (Akusha, Itsari, Mehweb, Muira, Tsudakhar); (3) Lak; (4) Lezgi (Archi, Rutul, Tabasaran); and (5) Turkic (Azerbaijani, Kumyk). The first four groups can be further clustered into a single group representing the Nakh-Daghestanian language family and compared to the languages belonging to the Turkic family. This parameter was taken into account to exclude diverging behaviors of speakers of different L1s. However, since the basic word order in the noun phrase is the same for all Nakh-Daghestanian and Turkic languages, we did not expect any remarkable difference.

Concerning the speaker's year of birth, we compared occurrences produced by younger and older speakers, that is, born before and after 1950. We chose 1950 as a watershed because it was starting from the end of the fifties that school education in highland Daghestan became more consistent for all children, including girls, and the first Russian teachers started to come to rural schools (Daniel & Dobrushina, 2013:188; Dobrushina, 2008:133). As a result, in this period the number of Russian speakers increased sharply: more than 90% of the inhabitants of highland Daghestan born after 1950 speak Russian (Dobrushina, Kozhukhar, & Moroz, 2019).

We checked the parameter "Education" by relying on the simple expectation that people with a higher degree of education might have higher proficiency in Russian, use it more actively, and be more aware of what is considered to be "standard."

The parameter "Gender" was included because in highland Daghestan women are usually less mobile, have a lower level of education and preserve a more traditional lifestyle. This implies that they probably use Russian less actively than men.

In [Figure 2](#), we show the general distribution of sociolinguistic parameters (excluding L1). Each point on the plot represents a speaker, the label indicates their place of birth, while the shape illustrates their level of education. The figures in brackets correspond to the total number of genitive noun phrases uttered by the speaker. The distribution of the points along the horizontal axis is determined by the speakers' year of birth, whereas their position along the vertical axis indicates the percentage of uses of prepositive genitives.

The most striking effect is seen for a number of speakers from Jangikent and Archib. We assume that this result might be explained simply in terms of idiosyncratic behaviors of individual speakers. Since there are no other visible correlations with place of birth and no particular expectations, this parameter was not included in the analysis.

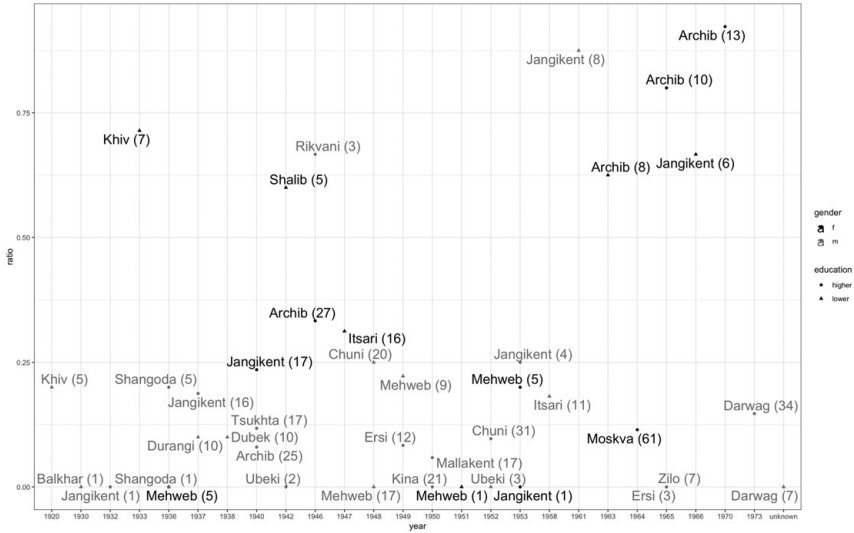


FIGURE 2. Ratio of prepositive genitives across speakers in the DagRus corpus by demographic features of the speakers: year of birth, gender and education (within parentheses are number of genitives produced).

Lexicosemantic features of the genitive modifier

In previous literature on genitive noun phrases, it was pointed out that the position of the genitive modifier along scales involving features such as animacy, referentiality, and individuation is relevant for the occurrence of different constructions (e.g., Koptjevskaja-Tamm, 2002:158; Lander, 2010:250-51). Specifically, nouns that are higher in such scales are more likely to behave in a special way due to their ability to function as better “anchors” (cf., Koptjevskaja-Tamm, 2002), that is, as elements serving as reference points (cf., Langacker, 1993) and helping the hearer identify the referent of the head.

Thus, we checked for such features as possibly relevant factors. First, we tagged our data according to the parameter “Lexical class,” which includes four classes distributed along the extended animacy scale: proper names (human referents) > kinship terms > other human referents > nonhuman referents.⁶ We annotated proper names and kinship terms separately because we expected that nouns placed on the left edge of the animacy scale would play a key role in the occurrence of left genitives. We were particularly interested in checking kinship terms, because in Daniel et al. (2010:85) it is mentioned that “left branching genitive chains that name relatives” are especially common in Daghestanian Russian. Moreover, most of the studies discussing word-order variation at the noun-phrase level (Friedman, 2003; Leisiö, 2000, 2001; Suleymanov, 2016; Sussex, 1993) point, more or less explicitly, to the occurrence of this type of variation in cases in which kinship semantics is involved. However, in most cases, no thorough discussion of this phenomenon is provided, and such observations have not been supported by statistical analyses.

Examples (5) to (8) illustrate each possible value of the parameter “Lexical class,” namely proper names (5), kinship terms (6), other human referents (7), and nonhuman referents (8).⁷

- 316 (5) *Vot èto Gusejnov Magomed vot Nadeždy*
 317 PTC this.N Gusejnov Magomed PTC **Nadežda.GEN**
 318 **Artemovny muž u nas žil**
 319 **Artemovna.GEN husband at 1 PL.GEN live.PST.SG.M**
 320 ‘Well, this Gusejnov Magomed, **Nadežda Artemovna’s husband**, lived at our
 321 place.’ [DagRus, L1 Lak]
- 322 (6) *Vot kak_raz-taki ja svoej tėti*
 323 PTC indeed 1SG REFL.F.GEN **aunt.GEN**
 324 **mogilu ne mogu najti**
 325 **grave.ACC NEG can.1SG find**
 326 ‘Well, indeed, I can’t find **my aunt’s grave**.’ [DagRus, L1 Mehweb Dargwa]
- 327 (7) *A eščë odna direktora našego žena*
 328 and also one.F **director.GEN our.M.GEN wife**
 329 ‘And one more, **our director’s wife**.’ [DagRus, L1 Mehweb Dargwa]
- 330 (8) *Vot moix let devuški*
 331 PTC **my.PL.GEN year.PL.GEN girl.PL**
 332 *uže nosili platki*
 333 already wear.PST.PL headscarf.PL
 334 ‘**Girls of my age** already wore headscarves.’ [DagRus, L1 Itsari Dargwa]

Second, we considered the parameter “Referentiality” and tagged the genitive nouns in our database based on the classification provided in (Padučeva, 1985:83-102) for Russian, that is, according to whether they denote a definite (9), indefinite (10), or nonspecific (11) referent; see also (Haspelmath, 1997:31-52) for a similar cross-linguistically oriented classification.

- 335 (9) *Vot èti kamni tože*
 336 PTC this.PL stone.PL also
 337 **ètogo mečeti kamni**
 338 **this.M.GEN mosque.GEN stone.PL**
 339 ‘These stones too, **the stones of this mosque** (...)’ [DagRus, L1 Azerbaijani]
- 340 (10) *V načale mnogie sosednix*
 341 in beginning.LOC numerous.PL **neighboring.PL.GEN**
 342 *sël byli učitelja*
 343 **village.PL.GEN COP.PST.PL teacher.PL**
 344 ‘At the beginning a lot (of them) were **teachers from the neighboring villages**.’ [DagRus, L1 Mehweb Dargwa]
- 345 (11) *No tam mne srazu efrejtora*
 346 but there 1SG.DAT immediately **Gefreiter.GEN**
 347 *zvanie dali*
 348 **rank.ACC give.PST.PL**
 349 ‘But there they immediately gave me **the Gefreiter rank**.’ [DagRus, L1
 350 Akusha Dargwa]

Finally, we tagged our data according to the parameter “Individuation,” that is, according to whether the noun in the genitive denotes a singular object, a plural object, a nonquantized mass, or an abstract entity (nonobject).

Lexicosemantic features of the head

As pointed out by Koptjevskaja-Tamm (2002:148), the semantics of the head noun usually determines the relation expressed within genitive noun phrases. Naccarato, Panova, and Stoyanova (2020) provided a detailed analysis of genitive noun phrases in Daghestanian Russian based on the semantic relation between the head noun and the genitive modifier (e.g., possessive, kinship, part-whole relations, and others, for a total of twenty-five relations). The analysis demonstrated that the most relevant opposition for predicting the occurrence of left genitives is kinship relations versus other relations.

Thus, provided that kinship relations are fully determined by the semantics of the head noun (kinship terms), we do not discuss semantic relations as a separate parameter, and we analyze the data according to the lexical class of the head noun, similarly to what was done for the genitive modifier. Specifically, we classified the head constituent into the following classes: kinship terms (12), other human referents (13), and nonhuman referents (14).

- (12) *Moej materi Aminatki*
 my.F.GEN mother.GEN Aminat.DIM.GEN
dvojurodnyj_brat byl predsedatelem
 cousin COP.PST.SG.M director.INS
 ‘My mother Aminat’s cousin was the director.’ [DagRus, L1 Mehweb Dargwa]

- (13) *Moego otca gosti byli oni*
 my.M.GEN father.GEN guests COP.PST.PL 3PL
 ‘They were my father’s guests.’ [DagRus, L1 Kumyk]

- (14) *Eë materi inja dali,*
 her.GEN mother.GEN name.ACC give.PST.PL
tože eë materi inja
 also her.GEN mother.GEN name.ACC
 ‘They gave her her mother’s name, also her mother’s name.’ [DagRus, L1 Kumyk]

Length of the genitive modifier and the head

As a possibly relevant factor affecting word order, we took into account the length of both the genitive modifier and the head constituent (one-word versus multiword); see (12) for examples of a multiword modifier (*moej materi Aminatki* ‘my mother Aminat’s’) and a multiword head (*dvojurodnyj brat* ‘cousin’). As observed in previous research, speakers tend to place heavier (i.e., longer) elements after lighter (i.e., shorter) ones (Hawkins, 1994:291-93; Wasow, 1997). Therefore, we might expect longer heads to be more frequently postposed to their genitive modifiers, and longer genitives to be more

frequently postposed to their heads. This expectation is also partly supported by the fact that one of the rare exceptions to the general tendency to left branching in some Daghestanian languages is constituted exactly by heavy genitive modifiers, which can sometimes be postposed to their heads (see, e.g., Kazenin [1996:150-151] on Godoberi).

Discourse-related factors

Deviations from the standard word order in the noun phrase might occur for discourse-related reasons. These might have to do with: (1) the givenness/newness of the information uttered (cf., the given-before-new principle); and (2) the specific information structure of a certain utterance (i.e., the need to topicalize or focalize certain elements in a phrase or sentence) (see, for example, Siewierska [1993] on both factors).

Checking the parameter “Givenness,” we were especially interested in cases including a given genitive and a new head, as in (15), because, according to the given-before-new principle, this kind of information structure should trigger the GEN+N order. All other possibilities (new genitive + given head, new genitive + new head, given genitive + given head) were expected to lead to the N+GEN order.

- (15) *i togda govorit Derbent podčinjalsja kaganatu*
 and then say.PRS.3SG Derbent obey.PST.SG.M Haganat.DAT
a kaganata centr byl Astraxani
 but Haganat.GEN center COP.PST.SG.M Astrakhan.LOC
 ‘And then, he says, Derbent was subordinate to the Haganat. And **the center of the Haganat** was in Astrakhan.’ [DagRus, L1 Azerbaijani]

Turning to information structure, we mentioned in the introductory section that in monolinguals’ Russian, prepositive genitives might appear due to the need to focus the genitive modifier, which takes phrasal accent in such uses, as in example (4) above. However, in this case we are probably dealing with a different syntactic structure. Focused genitives are not simply preposed to their heads but are dislocated to the left periphery of the clause (cf., [10] above). While annotating the data, we were not always able to differentiate between these two types of constructions. It is likely that some of the genitives annotated as left genitives in the DagRus sample are, in fact, left-dislocated and motivated by information structure, similarly to those found in monolinguals’ Spoken Russian.

NONSTANDARD WORD ORDER IN DAGHESTANIAN RUSSIAN:
 RESULTS AND SIGNIFICANCE OF THE PARAMETERS
 OBSERVED

Left genitives in Daghestanian Russian: descriptive statistics

The results of the annotation process are summarized in [Table 1](#). For each of the parameters considered, we calculated the percentage of left genitives on the total number of occurrences.

TABLE 1. *Parameters examined and the position of the genitive modifier*

	Parameter	Levels	GEN+N	N+GEN	% GEN+N
Sociolinguistics	L1	Nakh-Daghestanian	78	295	20.9%
		Turkic	25	84	22.9%
	Year of birth	< 1950	47	191	19.7%
		≥ 1950	56	188	22.9%
	Gender	F	63	119	34.6%
		M	40	260	13.3%
Education	Higher	62	275	18.4%	
	Lower	41	104	28.3%	
Lexicosemantic features of the genitive modifier	Lexical class	Proper name (human)	14	11	56%
		Kinship term	57	23	71.3%
		Human (other)	8	40	16.7%
		Nonhuman	24	305	7.3%
	Referentiality	Definite	89	222	28.6%
		Indefinite	1	17	5.6%
		Nonspecific	13	140	8.5%
	Individuation	Singular object	87	227	27.7%
		Plural object	5	44	10.2%
		Mass noun	1	9	1%
		Nonobject	10	99	9.2%
		Lexical class	Kinship term	68	23
Human (other)	16		132	10.8%	
Nonhuman	19		224	7.8%	
Length	Length of the genitive modifier	Multiword	52	142	26.8%
		One-word	51	237	17.7%
	Length of the head	Multiword	12	73	14.1%
		One-word	91	306	22.9%
Discourse-related factors	Givenness	New Head – Given Genitive	59	86	40.7%
		Other	44	293	13.1%
Total			103	379	21.4%

451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495

TABLE 2. *Binomial logistic regression: The optimal model predicting GEN+N*

	Estimate	<i>p</i> -value	<i>n</i>	%
(Intercept)	3.5057	<0.001		
Gender				
Male	reference level		300	62.2%
Female	-0.6120	NS	182	37.8%
Lexical class of the genitive modifier				
Nonhuman	reference level		329	68.3%
Human (other)	-0.5845	NS	48	9.9%
Proper	-0.9832	NS	25	5.2%
Kinship	-1.6019	<0.01	80	16.6%
Referentiality of the genitive modifier				
Nondefinite	reference level		171	35.5%
Definite	-0.6466	NS	311	64.5%
Lexical class of the head				
Nonkinship	reference level		391	81.1%
Kinship	-2.2855	<0.001	91	18.9%
Length of the genitive modifier				
One-word	reference level		397	82.4%
Multiword	-0.6624	<0.05	85	17.6%

As regards sociolinguistic parameters, we found a higher percentage of left genitives for female speakers and for speakers with a lower level of education (cf., Figure 2). Neither L1⁸ nor year of birth show a visible effect.

Turning to the lexicosemantic features of the genitive modifier, we see that all parameters seem to affect, to a greater or lesser extent, the word order in the noun phrase. As regards the lexical class of the genitive modifier, we may note that: (1) genitives denoting human referents are more likely to be preposed than those denoting nonhuman referents, and (2) the rate of left genitives is much higher for kinship terms and proper names than for other human referents. Left genitives are especially favored when the genitive is a kinship term, while for proper names the percentage is lower.⁹ The data on referentiality show that definite nouns are more frequently preposed to their heads compared to other nouns. Finally, regarding the parameter “Individuation,” we found that nouns denoting singular objects are more frequently preposed as compared to those denoting plural objects and masses. Nonobject nouns are also rarely preposed to the head. In the logistic model described below, we reduced the levels for the parameters “Referentiality” and “Individuation” to a binary outcome (i.e., definite versus nondefinite and singular versus nonsingular respectively).

As for the lexicosemantic features of the head, the data show that kinship terms (i.e., heads of noun phrases expressing a kinship relation) favor left genitives, whereas no notable difference is observed between other human referents and nonhuman referents. Therefore, in the logistic model, we reduced the possible outcomes for this parameter to kinship versus nonkinship.

541 The length of both the genitive modifier and the head affect, to some extent, the
 542 occurrence of a specific word order. Contrary to our expectations, multiword
 543 modifiers occur in the left position more frequently than one-word modifiers. As
 544 for the length of the head, left genitives more often occur in the presence of
 545 head constituents made up of a single word, while they occur less frequently
 546 when the head includes some modifiers.

547 Finally, concerning givenness, [Table 1](#) shows that the percentage of examples
 548 with left genitives is visibly higher when the genitive modifier is given and the
 549 head is new, which confirms our expectations.

550 Some of the above-mentioned factors are logically connected to each other, so
 551 we might expect that the frequency asymmetries shown in [Table 1](#) in fact represent
 552 the effect of a smaller set of factors. Heads with kinship semantics frequently co-
 553 occur with genitive modifiers also having kinship semantics (e.g., ‘my mother’s
 554 mother’). Referentiality, individuation, and givenness also interplay with the
 555 semantics of the genitive modifier: proper names and kinship terms tend to be
 556 definite, singular, and given. The parameter “Length” might be nonindependent
 557 as well. The high frequency of prepositive multiword genitive modifiers could
 558 be related to their co-occurrence with the class of kinship terms, which are often
 559 modified by possessive adjectives (*moej.GEN materi.GEN mat* ‘my mother’s
 560 mother’). Finally, education might correlate with gender.

561 *Binomial logistic regression*

562
 563 As the main statistical method, we used mixed-effects binomial logistic regression
 564 (Baayen, 2008:241-302; Levshina, 2015:171-98).¹⁰ All the parameters listed in
 565 [Table 1](#) were included in the model as fixed effects, that is, explanatory variables,
 566 whose effect has to be measured. We also included information about the speaker
 567 as a random effect to take into account possible idiosyncrasies. The results
 568 presented in [Table 2](#) include the optimal model that was obtained after removing
 569 the predictors that did not contribute to improving the model. Given that *p*-values
 570 alone do not constitute a sufficient reason to remove a variable from the model
 571 (Levshina, 2015:267), we resorted to AIC (Akaike Information Criterion)
 572 comparison, that is, we compared the AIC of the model including all predictors
 573 and of other models lacking one of the predictors, and removed (one at a time) the
 574 predictors whose presence in the model increased the AIC value (i.e., year of birth,
 575 givenness, L1, individuation of the genitive modifier, education, length of the head).

576 For each of the parameters included, in the column “Estimate” a negative
 577 coefficient boosts the chances of left genitives, whereas positive coefficients
 578 boost the chances of right genitives. The first estimate belongs to the intercept
 579 and gives us the estimated log odds of the outcome when all predictors are at
 580 their reference levels (Gender: Male, Lexical class of the genitive modifier:
 581 Nonhuman, Referentiality of the genitive modifier: Nondefinite, Lexical class of
 582 the head: Nonkinship, Length of the genitive modifier: One-word). The
 583 predictors including *p*-values lower than 0.05 are “Lexical class of the genitive
 584 modifier” and “Lexical class of the head,” both showing that kinship semantics
 585 boosts the chances of getting a left genitive, and “Length of the genitive

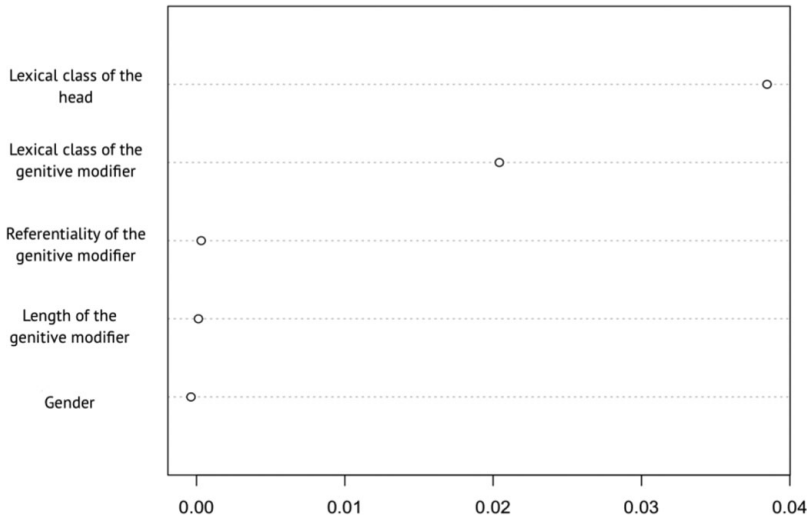


FIGURE 3. Random forest: Conditional importance of variables (number of trees = 1,000).

modifier,” showing that the chances of a left genitive to occur are significantly higher when the genitive has some modifiers.

Random forest

For the final set of predictors, in addition to the logistic regression analysis, we implemented a random forest method,¹¹ which is particularly useful when the number of observations is relatively small, but the number of predictors is large (Levshina, 2015:291-300; Tagliamonte & Baayen, 2012), as in our case. Random forests (and conditional inference trees, out of which random forests are grown) are based on a recursive binary splitting of the data, which partitions the input data into increasingly homogeneous subsets, and help determine whether a variable is a useful predictor. The importance measure of each variable is averaged over many conditional trees. The model represented in Figure 3 includes all the predictors from the optimal logistic regression model in Table 2. The predictors in the rightmost part of the chart are those whose impact in the model proved to be stronger.

According to the results shown in Figure 3, the predictor that best discriminates the data is the lexical class of the head, followed by the lexical class of the genitive modifier. The other predictors are found in the leftmost part of the chart, which is a sign of their lower strength.

In general, these results are comparable to those of the logistic regression analysis, as the same variables were found to be the most relevant according to both analyses. Overall statistical computations confirmed our expectation that left genitives in Daghestanian Russian are semantically determined and,

TABLE 3. *Monolinguals' Russian corpora: size, extracted and selected occurrences*

Corpus	Size	Extracted occurrences	Selected occurrences
RNC, Memories	~ 54,000	1,462	151
RNC, Conversation	~ 263,000	8,083	401
Zvenigorod	~ 68,000	2,018	123
Total	~ 385,000	11,563	675

precisely, that kinship semantics is a crucial factor in these respects, further supporting the idea put forth in the description of Daghestanian Russian by Daniel et al. (2010:85-86). More precisely, we showed that the lexical class of both the head and the genitive modifier individually contribute to the choice of a specific word order. We also found that other factors make their own smaller contribution, too. These are referentiality and length of the genitive modifier, and the speaker's gender.

NOUN PHRASES WITH A GENITIVE MODIFIER IN MONOLINGUALS' SPEECH

To understand the extent to which nonstandard word order in Daghestanian Russian is determined by language contact, we looked at the production of noun phrases with a genitive modifier in the speech of Russian monolinguals and compared these data with our Daghestanian data.

The monolinguals' Russian data were retrieved from three oral text collections, which are more or less comparable to that of the DagRus corpus in terms of size, text types, and genres (Table 3). Two collections come from the Spoken Subcorpus of the RNC (see <http://ruscorpora.ru/old/en/search-spoken.html>). These are a collection of oral texts produced by Muscovites about their memories (*Vospominanija o prošlom*, cf., Kitajgorodskaja & Rozanova, 1999) and a manually compiled collection of oral texts tagged in the RNC as "Conversation." The third collection is the Corpus of Russian spoken in Zvenigorod (Moscow Oblast), which was created ad-hoc for this study (Panova, 2020, see <http://lingconlab.ru/zvenigorod/>).¹² We selected relevant occurrences from these texts following the same procedure as for the Daghestanian data, and we were left with 675 examples, which is comparable to the Daghestanian sample.

We annotated the resulting sample according to the parameters that proved to be relevant for the Daghestanian data.¹³ As we expected, the total number of left genitives in the monolinguals' Russian sample is much smaller than in the DagRus sample: only nineteen out of 675 examples (2.8%) contain prepositional genitives, as in (16).¹⁴

- (16) *U nas=to vot muža sestra*
at 1PL.GEN=PTC PTC husband.GEN sister

'And here is my husband's sister...' [RNC, Memories]

TABLE 4. *Monolinguals' Russian: position of the genitive modifier according to parameters relevant for Daghestanian Russian*

Parameter	Levels	GEN+N	N+GEN	% GEN+N
Lexical class of the genitive modifier	Proper	9	61	12.9%
	Kinship	3	31	8.8%
	Human (other)	1	58	1.7%
Referentiality of the genitive modifier	Nonhuman	6	506	1.2%
	Definite	12	329	3.5%
Lexical class of the head	Nondefinite	7	327	2.1%
	Kinship	8	38	17.4%
Length of the genitive modifier	Nonkinship	11	618	1.8%
	One-word	10	425	2.3%
	Multiword	9	231	3.8%

We did not conduct the same statistical analysis as for the DagRus data due to the small number of left genitives found in the sample. However, the preliminary results summarized in Table 4 show that the attested frequency asymmetries are similar to those observed for the DagRus data.

Although the length and referentiality of the genitive modifier do not show any effect, the lexical class of both the head and the genitive modifier significantly affect the choice of a specific word order. Heads with kinship semantics are more likely to take prepositional genitive modifiers compared to other types of heads.¹⁵ Genitive modifiers constituted by kinship terms and human proper names are also more likely to be preposed.¹⁶ Thus, the strongest predictors of left genitives are the same in both varieties.

SUMMARY AND DISCUSSION

Summary of the results obtained

In this paper, we showed that genitive modifiers in Daghestanian Russian tend to be preposed to their heads, which gives rise to the same word-order pattern as in the speakers' L1s. Remarkably, this tendency does not equally affect all types of genitive noun phrases. We showed that the position of the genitive modifier is determined by sociolinguistic parameters, lexicosemantic and formal features of the noun phrase. The results are summarized in Table 5.

Left genitives tend to be more frequent in texts produced by female speakers and speakers with a lower level of education. With respect to linguistic factors, preposition of the genitive correlates with lexicosemantic features of both the head and the genitive modifier. Specifically, left genitives and their heads frequently belong to the class of kinship terms. Left genitives are often definite and, to a large extent, refer to countable singular objects. In addition, they are frequently multiword. By contrast, the head of a prepositional genitive is usually a one-word constituent (i.e., it has no other modifiers). Preposition of the genitive

TABLE 5. *Daghestanian Russian: summary of the results*

Type of parameter	Parameter	Value(s) associated with left genitives	Statistically significant
Sociolinguistic	L1 (language family)	No visible correlation	no
	Year of birth	No visible correlation	no
	Gender	Female	no* ¹⁷
Lexicosemantic	Education	Lower	no
	Lexical class of the genitive modifier	Proper	no
		Kinship	yes
	Referentiality of the genitive modifier	Human (other)	no
		Definite	no*
Length	Individuation of the genitive modifier	Singular	no
	Lexical class of the head	Kinship	yes
	Length of the genitive modifier	Multiword	yes
Discourse-related	Length of the head	One-word	no
	Givenness	New Head-Given Genitive	no

also correlates with givenness: left genitives are more frequent in noun phrases in which the referent of the genitive modifier is given and the referent of the head is new. Some of these factors seem to be logically connected to each other. Statistical analysis reduces the list of relevant factors to a smaller number of parameters that are enough to predict the position of the genitive modifier (see the rightmost column in Table 5). These parameters affect word order in the noun phrase to different extents, and their impact could be represented in the hierarchy shown in (17).

- (17) lexical class of the head > lexical class of the genitive modifier > length of the genitive modifier > referentiality of the genitive modifier > speaker's gender

The comparison with monolinguals' Russian showed that left genitives in Daghestanian Russian are significantly more widespread than in monolinguals' Russian (21.4% versus 2.8%). However, the most relevant factors for the occurrence of left genitives in Daghestanian Russian appeared to be relevant for monolinguals' Russian too.

Word-order variation within the noun phrase: a more general perspective

The results of our study generally agree with findings concerning other varieties of contact-influenced Russian. Suleymanov (2016:18-19), in his discussion of left-branching genitive constructions in Baku Russian (which is geographically close to Daghestanian Russian), provided several examples of this phenomenon, and in all of them either the head or the genitive modifier is a kinship term. In

766 analyzing possessive adjectives and genitives in two varieties of Finland Russian,
 767 Leisiö (2000, 2001) explicitly says that the order “modifier-head” is common when
 768 the modifier is a “specific personal possessor, usually a kinship term or proper
 769 name” (Leisiö, 2000:308) and appeals to such notions as familiarity,
 770 relationality, givenness, as well as to the empathy hierarchy (DeLancey, 1981).

771 Comparable factors determine synchronic splits and diachronic shifts in the
 772 marking of genitive noun phrases in noncontact settings too. For example, a
 773 methodologically similar study (Szmrecsanyi, 2013) on the competition between
 774 the *s*-genitive and the *of*-genitive in English concludes that language-internal
 775 conditioning factors are the following: genitive relation (ownership versus other
 776 relations), possessor animacy, possessum length and possessor thematicity (cf.,
 777 also Szmrecsanyi, Biber, Egbert, & Franco, 2016). In her study on Ancient
 778 Greek, Viti (2008) showed that word-order variation in genitive noun phrases is
 779 regulated by such parameters as semantic relation between the head and the
 780 genitive (kinship versus part-whole, material and content, measure), as well as
 781 animacy, humanness, individuation, and topicality of the genitive modifier.

782
 783 *Word-order variation in Daghestanian Russian: possible*
 784 *explanations*
 785

786 Some of the parameters that turned out to be relevant for word-order variation in
 787 Daghestanian Russian can be explained by appealing directly to word-order
 788 processing considerations.

789 As we saw, the most relevant factor triggering the occurrence of left genitives is
 790 kinship semantics of the head of the noun phrase and, consequently, the semantic
 791 relation between the head and the genitive modifier. A possible explanation lies in
 792 the processing easing that left-branching constructions can bring about when
 793 discussing kinship relations. A head with kinship semantics is “anchored” by
 794 relating it to a genitive modifier that serves as a reference point for identifying
 795 its referent (cf., Hawkins [1991:411] on “anchoring” terminology). It is
 796 compelling to believe that it might be easier for the speaker to start with the
 797 anchor, that is, the item that is closest to him/her and, possibly, to his/her
 798 interlocutor. Within this view, word order would iconically reflect the
 799 proximity/familiarity of the speaker and the addressee with the referents
 800 mentioned in the noun phrase (i.e., “left” = “more familiar,” see also Leisiö
 801 [2001:106-107]). This is especially relevant for complex chains of kinship
 802 relations with several genitive modifiers (such as *moego.GEN brata.GEN ženy.GEN*
 803 *mat* ‘my brother’s wife’s mother’).

804 The parameter “Givenness” can be explained in a similar way. The fact that
 805 given genitive modifiers tend to be preposed to new heads seems to reflect the
 806 general given-before-new principle, which can help the addressee understand the
 807 flow of information coming next.

808 The last parameter specifically associated with word order is the length of the
 809 genitive modifier. The general cross-linguistic tendency is that heavy modifiers
 810 are more prone to postposition. Our data do not follow this principle: in

811 Daghestanian Russian multiword (heavy) genitive modifiers tend to be preposed
 812 instead. It appears that the left multiword genitive modifiers attested in our
 813 sample do not form a homogenous class, and their frequent preposition may
 814 have different motivations. Typical examples are genitive chains (*materi.GEN*
 815 *brata.GEN* ‘of (my) mother’s brother’), genitives with possessive pronouns (*moej.*
 816 *GEN sestry.GEN* ‘of my sister’) and deictic markers (*ëtogo.GEN čeloveka.GEN* ‘of
 817 this person’), genitive modifiers expressing age characteristics (*požilogo.GEN*
 818 *vozrasta.GEN* ‘of old age’), proper names (*Šaxmara.GEN Omaroviča.GEN* ‘of
 819 Shahmar Omarovich’). Some of these uses might be motivated by the factors of
 820 definiteness and kinship semantics, while others need further research (see also
 821 endnote 2 on preposed multiword genitive modifiers denoting qualities in
 822 monolinguals’ Russian).

823 The other parameters, that is, the lexical class and referentiality of the genitive
 824 modifier, cannot be directly explained in terms of word-order processing. They are
 825 often mentioned in studies discussing the mechanisms lying behind similar splits
 826 (albeit not strictly related to word order) in noncontact situations. These can be
 827 specific mechanisms responsible for splits in possessive noun phrases (cf.,
 828 Koptjevskaja-Tamm, 2002), or more general ones, regulating different types of
 829 splits in argument encoding. The cluster of these parameters is associated with
 830 such general notions as animacy, agentivity, individuation, empathy, topicality,
 831 in proposed hierarchies such as the “empathy hierarchy” (DeLancey, 1981; Kuno
 832 & Kaburaki, 1977), “hierarchy of inherent lexical content” (Silverstein, 1976),
 833 “activity scale” (Moravcsik, 1978), “extended animacy hierarchy” (Croft,
 834 2003:130), “individuation properties” (Hopper & Thompson, 1980), “topicality
 835 hierarchy” (Lander, 2010) (see also Filimonova [2005] and Haude & Witzlack-
 836 Makarevich [2016] for an overview of such hierarchies). The results of our study
 837 show that most of the features characterizing left genitive modifiers occupy
 838 higher positions along such hierarchies.

839 Word order is indirectly connected to contexts associated with higher/lower
 840 position on such semantic hierarchies through the notion of anchoring, which
 841 was already employed above to explain the first cluster of parameters. Anchors,
 842 which tend to be preposed to their heads, occupy high positions along the
 843 hierarchies (Koptjevskaja-Tamm, 2002).

844 However, if anchoring was the only mechanism regulating word-order choice,
 845 we would expect a strong effect caused by the presence of relational nouns—
 846 other than kinship terms—as heads (e.g., body parts, group members, etc.).
 847 Given their property of expressing a relation with some other entity, relational
 848 nouns grammatically require the expression of this entity as an anchor for their
 849 understanding. However, the data do not demonstrate any substantial difference
 850 between relational nouns and nonrelational nouns.¹⁸ Moreover, a closer look at
 851 the lexical class of each head noun in our corpus demonstrates that the class of
 852 kinship terms is actually the only group of relational nouns strongly correlating
 853 with the occurrence of left genitives.¹⁹

854 In a contact situation, the motivations behind word-order variation discussed
 855 above might be enhanced by language-external factors. We suppose that, in the

environment of highland Daghestan to which our data belong, kinship semantics pertains to the topics usually discussed in L1 and not in Russian. Moreover, lexical items belonging to this domain are allegedly learned at an early stage of L1 acquisition. Therefore, L1 influence in this domain might be stronger than in other contexts (see, e.g., Aikhenvald [2007:30] on the “communicative salience” of kinship terms, which makes this lexical group more prone to contact-induced changes).

CONCLUSION

In this paper, we presented a quantitative corpus-based investigation of noun phrases with a genitive modifier in the contact variety of Russian spoken in Daghestan. Such phrases are characterized by a frequent GEN+N word order, which is not typical of monolinguals’ (i.e., noncontact) varieties of Russian.

The results of our study contribute to the more general discussion on the contact versus noncontact nature of grammatical changes observed in contact settings. As discussed in the introductory section, in recent literature it has been pointed out that contact is often invoked to explain changes that rather result from a more complex combination of factors. In some cases, contact merely reinforces certain language-internal tendencies that are better explained in terms of more general underlying functional (cognitive) mechanisms. Word-order variation in Daghestanian Russian has proved to be another example of such nontrivial cases for which contact influence alone is not a sufficient explanation.

In this paper, we showed that contact influence manifests itself in the unexpectedly high frequency of the nonstandard word-order pattern GEN+N attested in Daghestanian Russian, but not in the factors regulating its occurrence. Word-order variation in the Russian speech of Daghestanian bilinguals is determined by a set of parameters that have been recognized as cross-linguistically relevant for the occurrence of similar morphosyntactic splits—not necessarily involving word order—in noncontact situations, as discussed above. Some of these factors proved to be relevant for the choice of the GEN+N word-order pattern in monolinguals’ Russian too, although in monolinguals’ varieties this phenomenon is extremely infrequent. At the same time, it is important to stress that in the languages of Daghestan, word order in noun phrases is not regulated by the factors discussed in our study.

Thus, the constraints regulating word order in Daghestanian Russian can indeed be formulated without appealing to language contact, although contact might be seen as a factor enhancing the observed variation.

As a follow-up to the analysis presented in this paper, it would be interesting to consider left-branching constructions in Daghestanian Russian from the point of view of language prescriptivism, that is, to check the hypothesis that the expansion of the GEN+N word-order pattern might also be connected to the fact that Daghestanian Russian is relatively free from the pressure of

901 prescriptive norms. If so, this process might be attributed to language “natural
902 development” in the sense of Chambers (2004), and the choice of the
903 GEN+N pattern in certain pragmatically salient positions might be considered
904 as a “vernacular universal” (Filppula, Klemola, & Paulasto, 2009). The latter
905 term reflects the more general hypothesis that the grammar of such
906 nonnormalized varieties is more affected by universal cognitive-based
907 principles than that of a normalized standard language.

908 Evidence supporting or rejecting this hypothesis could be obtained through the
909 comparison of our Daghestanian data with the data of monolinguals’ Russian
910 dialects, which are also free from prescriptive norms but are not affected by
911 language contact (at least at the synchronic stage). According to our preliminary
912 observations, the GEN+N pattern is indeed quite widespread in monolinguals’
913 Russian dialects: in the available corpora of dialectal Russian,²⁰ the genitive
914 modifier is proposed to its head on average in 18% of the cases.

915 As shown in Thomason (2009), contact-induced change is often hardly
916 distinguishable from vernacular universals, and the case of word-order variation
917 discussed in this paper provides a good illustration of this claim. A deeper
918 investigation of the factors conditioning the occurrence of nonstandard word
919 order in monolinguals’ dialectal varieties of Russian, as well as in bilinguals’
920 varieties of Russian other than those spoken in Daghestan, might be the topic of
921 future research. This could help better understand the complicated interaction of
922 contact influence and language natural development in the field of word-order
923 variation.

924 925 926 ACKNOWLEDGMENTS

927
928 A preliminary version of the paper was presented at the “Various Dimensions of
929 Contrastive Studies” conference (University of Silesia in Katowice, October 21–
930 22, 2019). Subsequent versions of the paper were presented in Moscow during
931 seminars organized at the National Research University Higher School of
932 Economics and at the Moscow State University. We would like to thank our
933 colleagues at the Linguistic Convergence Laboratory (National Research
934 University Higher School of Economics) for useful discussion and valuable
935 advice, in particular Nina Dobrushina, Michael Daniel, and Yury Lander for
936 constant consultations during the preparation of the paper, Ilya Schurov, George
937 Moroz, Ezequiel Koile, Olga Lyashevskaya, and Anna Klezovich for their help
938 with statistical computations. We would also like to thank Nina Dobrushina,
939 Peter Arkadiev, and Samira Verhees, who read a preliminary version of the
940 paper and provided us with insightful suggestions. Irina Khomchenkova and
941 Polina Pleshak contributed to the annotation and discussion of the
942 monolinguals’ Russian data. Finally, we would like to thank four anonymous
943 reviewers for *Language Variation and Change*, whose comments significantly
944 contributed to improving the paper. The article was prepared within the
945 framework of the HSE University Basic Research Program.

NOTES

1. Russian glosses are simplified: we do not indicate morphological boundaries, gender is not glossed for nouns, singular number is not glossed for nouns and adjectives, nominative case is never glossed. The abbreviation “RNC” stands for Russian National Corpus (see <http://ruscorpora.ru/new/>), whereas “Memories” stands for a collection included in the spoken subcorpus of the RNC. The abbreviation “DagRus” refers to the Corpus of Russian spoken in Daghestan.

2. In addition, left genitives are regularly used in some collocations (cf., *gvardii.GEN major* ‘guard’s major’), and are deemed acceptable in noun phrases in which multiword genitive modifiers denote some quality of their heads (*zelenogo.GEN cveta.GEN oboi* ‘green wallpaper,’ *moego.GEN vozrasta.GEN ljudi* ‘people of my age’) (Graščenkov, 2016; Švedova, 1980:204-5). Such cases are rare and more typical of written texts.

3. The map was created with the R package “lingtypology” (Moroz, 2017).

4. In some languages of Daghestan, the order N+GEN is found with numerals and is acceptable or even neutral in quantitative expressions (see Sumbatova & Lander [2014:204] on Tanty Dargwa [Dargwa, Nakh-Daghestanian]).

5. All the examples excluded from the DagRus sample feature the N+GEN word-order pattern.

6. Nouns denoting animals were not treated as a separate class because they are very infrequent in our data (three out of 482, i.e., 0.6%; all of them are right genitives).

7. In (5)-(16), we deliberately report examples with left genitives. However, it should be noted that examples with right genitives were also found (see Table 1 for the distribution of left versus right genitives).

8. For reasons of space, for the parameter “L1” we only reported the figures comparing the general results for the two linguistic families (Nakh-Daghestanian and Turkic). Note, however, that the comparison of single languages, language groups, and families does not reveal any substantial difference in the occurrence of left genitives (cf., Naccarato et al., 2020:150).

9. Note, however, that a small part of proper names might not have been identified by the parser as genitives as, for example, feminine names ending in a consonant, which are not inflected. So this group might be slightly underrepresented in our sample.

10. We used the function *glmer()* from the R package “lme4.” The full R code with the statistical analysis and the dataset are available at https://osf.io/d2cgg/?view_only=cde1da424fb54e56933109f9e78dbc24.

11. We used the function *cforest()* from the R package “party.”

12. To make this collection comparable to the DagRus Corpus in terms of the topics discussed, we intentionally followed the same procedure of conducting interviews that was used in Daghestan.

13. Speaker’s gender was not included in the annotation because this information was only partially available.

14. One might assume that left genitives in Daghestanian Russian correspond functionally to possessive adjectives (*mamin.POSS brat* ‘(my) mother’s brother’) and/or to the construction ‘*u* + GEN’ (*u sestry.GEN muž umer* ‘(my) sister’s husband died’) in monolinguals’ Russian. However, we failed to find any convincing evidence based on our data, since both constructions are attested in Daghestanian Russian, too.

15. Two-tailed Fisher’s exact test: $p < 0.0001$.

16. Two-tailed Fisher’s exact test: $p < 0.0001$ (proper and kinship versus human (other) and nonhuman).

17. We indicate with an asterisk those values that are present in the optimal logistic regression model but whose *p*-values turned out to be nonsignificant (see Table 2).

18. We found 22.8% (77 out of 338) left genitives for relational nouns versus 18.1% (twenty-six out of 144) for nonrelational nouns in our sample; the difference is not statistically significant (two-tailed Fisher’s exact test: $p = 0.2757$).

19. We found 74.7% left genitives for kinship terms (sixty-eight out of ninety-one) versus 3.6% for other types of relational nouns (nine out of 247), the difference is statistically significant (two-tailed exact Fisher’s test: $p < 0.0001$).

20. Ustyia dialect (Arkhangelsk Oblast), see <http://www.parasolcorpus.org/dagrus/>; Opočka dialect (Pskov Oblast), see <https://lingconlab.ru/opochka/>; Malinino dialect (Lipetsk Oblast), see <https://lingconlab.ru/malinino/>; Rogovatká dialect (Belgorod Oblast), see <http://www.parasolcorpus.org/Rogovatká/>; Spiridonova Buda dialect (Bryansk Oblast), see <http://lingconlab.ru/spiridonovabuda/>; Tambov Oblast dialects, see <http://ruscorpora.ru/old/search-dialect.html>.

REFERENCES

- 991
992 Aikhenvald, Alexandra Y. (2007). Grammars in Contact: A Cross-Linguistic Perspective. In A. Y.
993 Aikhenvald & R. M. W. Dixon (Eds.), *Grammars in Contact. A Cross-Linguistic Typology*.
994 Oxford: Oxford University Press. 1–66.
- 995 Aikhenvald, Alexandra Y. & Dixon, Robert M. W. (Eds.). (2007). *Grammars in Contact. A Cross-*
996 *Linguistic Typology*. Oxford: Oxford University Press.
- 997 Baayen, Harald R. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*.
998 Cambridge: Cambridge University Press.
- 999 Bickel, Balthasar. (2017). Areas and Universals. In R. Hickey (Ed.), *The Cambridge Handbook of Areal*
1000 *Linguistics*. Cambridge: Cambridge University Press. 40–54.
- 1001 Campbell, Lyle. (1997). Typological and areal issues in reconstruction. In J. Fisiak (Ed.), *Linguistic*
1002 *reconstruction and typology*. Berlin: Mouton de Gruyter. 45–72.
- 1003 Chambers, J. K. (2004). Dynamic typology and vernacular universals. In B. Kortmann (Ed.),
1004 *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*.
1005 Berlin/New York: Mouton de Gruyter. 127–45.
- 1006 Croft, William. (2003). *Typology and universals*. Cambridge: Cambridge University Press.
- 1007 Dahl, Östen. (2001). Principles of areal typology. In M. Haspelmath, E. König, W. Oesterreicher, & W.
1008 Raible (Ed.), *Language Typology and Language Universals. An International Handbook. Volume 2*.
1009 Berlin/New York: De Gruyter Mouton. 1456–71.
- 1010 Daniel, Michael & Dobrushina, Nina. (2010). Novye russkie [New Russians]. *Voprosy russkogo*
1011 *jazykoznanija* 13:141–58.
- 1012 ———. (2013). Russkij jazyk v Dagestane: problemy jazykovej interferencii [Russian in Daghestan:
1013 problems of language interference]. In *Computational Linguistics and Intellectual Technologies.*
1014 *Papers from the Annual International Conference "Dialogue" (2013). Issue 12*. Moscow: RSUH.
1015 186–99.
- 1016 Daniel, Michael, Dobrushina, Nina, & Knyazev, Sergey. (2010). Highlander's Russian: Case Study in
1017 Bilingualism and Language Interference in Central Daghestan. *Instrumentarium of Linguistics:*
1018 *Sociolinguistic Approach to Non-Standard Russian. Slavica Helsingiensia* 40:65–93.
- 1019 DeLancey, Scott. (1981). An interpretation of split ergativity and related patterns. *Language* 57:626–57.
- 1020 Dobrushina, Nina. (2008). Tradicii i innovacii v kul'ture gornogo Dagestana: iz istorii sovetskoj školy
1021 [Traditions and innovations in the culture of highland Daghestan: from the history of the Soviet
1022 school]. *Voprosy obrazovanija* 3:119–44.
- 1023 Dobrushina, Nina, Daniel, Michael, von Waldenfels, Ruprecht, Maisak, Timur, & Panova, Anastasia.
1024 (2018). *Corpus of Russian spoken in Daghestan*. Moscow: Linguistic Convergence Laboratory,
1025 NRU HSE. (Available online at <http://www.parasolcorpus.org/dagrus/>, accessed on 21.08.2018.)
- 1026 Dobrushina, Nina, Kozhukhar, Aleksandra, & Moroz, George. (2019). Gendered Multilingualism in
1027 highland Daghestan: story of a loss. *Journal of Multilingual and Multicultural Development* 40
1028 (2):115–32.
- 1029 Dobrushina, Nina & Kultepina, Olga. (2021). The rise of a lingua franca: The case of Russian in
1030 Daghestan. *International Journal of Bilingualism* 25(1):338–58.
- 1031 Filimonova, Elena. (2005). The noun phrase hierarchy and relational marking: Problems and
1032 counterevidence. *Linguistic Typology* 9:77–113.
- 1033 Filppula, Markku, Klemola, Juhani, & Paulasto, Heli (Eds.). (2009). *Vernacular universals and*
1034 *language contacts: Evidence from varieties of English and beyond*. London: Routledge.
- 1035 Friedman, Victor A. (2003). *Turkish in Macedonia and Beyond: Studies in Contact, Typology and other*
1036 *Phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz.
- 1037 Graščenkov, Pavel V. (2016). Sintaksičeskaja i semantičeskaja tipologija genitiva: genitiv kačestva kak
1038 nositel' ad'ektivnogo priznaka (na materiale russkogo jazyka) [Syntactic and semantic typology of
1039 the genitive: the genitive of quality as carrier of adjectival features (based on Russian material)]. In
1040 E. A. Ljutikova & A. V. Zimmerling (Eds.), *Arhitektura klauzy v parametričeskix modeljax*
1041 [Clause architecture in parametric models]. Moscow: Jazyki slavjanskix kul'tur. 444–67.
- 1042 Haspelmath, Martin. (1997). *Indefinite pronouns*. Oxford: Oxford University Press.
- 1043 Haude, Katharina & Witzlack-Makarevich, Alena. (2016). Referential hierarchies and alignment: An
1044 overview. *Linguistics* 54(4):433–41.
- 1045 Hawkins, John A. (1991). On (in)definite articles: implicatures and (un)grammaticality prediction.
1046 *Journal of linguistics* 27(2):405–42.
- 1047 ———. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University
1048 Press.
- 1049
- 1050

- 1036 Heine, Bernd. (2008). Contact-induced word order change without word order change. In P. Siemund &
 1037 N. Kintana (Eds.), *Language Contact and Contact Languages*. Amsterdam/Philadelphia: John
 Benjamins. 33–60.
- 1038 Heine, Bernd & Nomachi, Motoki. (2013). Contact-induced replication. Some diagnostics. In M.
 1039 Robbeets & H. Cuyckens (Eds.), *Shared Grammaticalization: With special focus on the*
 1040 *Trans Eurasian languages*. Amsterdam/Philadelphia: John Benjamins. 67–100.
- 1041 Hickey, Raymond (Ed.). (2017). *The Cambridge Handbook of Areal Linguistics*. Cambridge:
 Cambridge University Press.
- 1042 Hopper, Paul J. & Thompson, Sandra A. (1980). Transitivity in grammar and discourse. *Language*
 1043 56:251–99.
- 1044 Kazenin, Konstantin. 1996. Word order in noun phrase. In A. E. Kibrik, S. G. Tatevosov, & A.
 Eulenberg (Eds.), *Godoberi*. München/Newcastle: Lincom Europa. 148–54.
- 1045 Kitajgorodskaja, Margarita V. & Rozanova, Nina N. (1999). *Reč' moskvičej. Kommunikativno-*
 1046 *kul'turoložičeskij aspekt* [The speech of Muscovites. Communicative and cultural aspects].
 Moscow: Russkie slovari.
- 1047 Koptjevskaja-Tamm, Maria. (2002). Adnominal possession in the European languages: form and
 1048 function. *Sprachtypologie und Universalienforschung* 55:141–72.
- 1049 Kortmann, Bernd. (Ed.). (2004). *Dialectology Meets Typology. Dialect Grammar from a Cross-*
Linguistic Perspective. Berlin/New York: De Gruyter Mouton.
- 1050 Kuno, Susumu & Kaburaki, Etsuko. (1977). Empathy and Syntax. *Linguistic Inquiry* 8(4):627–72.
- 1051 Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- 1052 Lander, Yury A. (2010). Dialectics of adnominal modifiers: On concord and incorporation in nominal
 1053 phrases. In F. Floricic (Ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis*
Creissels. Lyon: ENS éditions. 247–70.
- 1054 Langacker, Ronald W. (1993). Reference-point constructions. *Cognitive Linguistics* 4(1):1–38.
- 1055 Leisiö, Larisa. (2000). The word order in genitive constructions in a diaspora Russian. *International*
Journal of Bilingualism 4(3):301–25.
- 1056 ———. (2001). *Morphosyntactic Convergence and Integration in Finland Russian*. Doctoral
 1057 dissertation, University of Tampere.
- 1058 Levshina, Natalia. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*.
 Amsterdam/Philadelphia: John Benjamins.
- 1059 Matras, Yaron, McMahon, April & Nigel, Vincent. (Eds.). (2006). *Linguistic Areas: Convergence in*
 1060 *Historical and Typological Perspective*. London: Palgrave.
- 1061 Moravcsik, Edith A. (1978). On the distribution of ergative and accusative patterns. *Lingua* 45:233–79.
- 1062 Moroz, George. (2017). *_lingtypology: easy mapping for Linguistic Typology_*. [https://CRAN.R-](https://CRAN.R-project.org/package=lingtypology)
[project.org/package=lingtypology](https://CRAN.R-project.org/package=lingtypology).
- 1063 Mougeon, Raymond, Nadasdi, Terry, & Rehner, Katherine. (2005). Contact-induced linguistic
 1064 innovations on the continuum of language use: The case of French in Ontario. *Bilingualism:*
Language and Cognition 8(2):99–115.
- 1065 Muysken, Pieter. (2008). *From Linguistic Areas to Areal Linguistics*. Amsterdam/Philadelphia: John
 1066 Benjamins.
- 1067 Naccarato, Chiara, Panova, Anastasia, & Stoynova, Natalia. (2020). Nestandartnyj porjadok slov v
 1068 dagestanskom russkom: imennye gruppy s genitivom [Non-standard word order in Daghestanian
 1069 Russian: noun phrases with the genitive]. In *Jazyki narodov Rossii v kontakte s russkim jazykom.*
Trudy Instituta russkogo jazyka im. V. V. Vinogradova 4:146–67.
- 1070 Nichols, Johanna. (1992). *Linguistic diversity in space and time*. Chicago/London: The University of
 Chicago Press.
- 1071 Padučeva, Elena V. (1985). *Vyskazyvanie i ego sootnesenost' s dejstvitel'nost'ju. Referencial'nye*
 1072 *aspekty semantiki mestoimenij* [The utterance and its relationship to reality: Referential aspects of
 1073 the semantics of pronouns]. Moscow: LKI.
- 1074 Panova, Anastasia. (2020). *Corpus of Russian spoken in Zvenigorod*. Moscow: Linguistic Convergence
 1075 Laboratory, HSE University. (Available online at <http://lingconlab.ru/zvenigorod/>,
 1076 accessed on 02.03.2021.)
- 1077 Poplack, Shana. (2021). A variationist perspective on language contact. In E. Adamou & Y. Matras
 1078 (Eds.), *The Routledge Handbook of Language Contact*. London/New York: Routledge. 46–62.
- 1079 Poplack, Shana & Levey, Stephen. (2010). Contact-induced grammatical change: A cautionary tale. In P.
 1080 Auer & J. E. Schmidt (Eds.), *Language and space: An international handbook of linguistic variation.*
Volume 1: Theories and Methods. Berlin/New York: De Gruyter Mouton. 391–419.

- 1081 Poplack, Shana, Zentz, Lauren, & Dion, Nathalie. (2012). Phrase-final prepositions in Quebec French:
 1082 An empirical study of contact, code-switching and resistance to convergence. *Bilingualism:
 1083 Language and Cognition* 15(2):203–25.
- 1084 Röthlisberger, Melanie & Szmrecsanyi, Benedikt. (2019). Dialect typology: Recent advances. In S.
 1085 Brunn & R. Kehrein (Eds.), *The Handbook of the Changing World Language Map*. Cham:
 1086 Springer. 1–26.
- 1087 Siewierska, Anna. (1993). On the interplay of factors in the determination of word order. In J. Jacobs, A.
 1088 von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: An International Handbook of
 1089 Contemporary Research*. Berlin: Walter de Gruyter. 826–46.
- 1090 Silverstein, Michael. (1976). Hierarchy of features and ergativity. In R.M.W. Dixon (Ed.), *Grammatical
 1091 categories in Australian languages*. New Jersey: Humanities Press. 112–71.
- 1092 Suleymanov, Murad. (2016). Baku Russian: Colonial heritage in a post-colonial era. In D. Schmidt-
 1093 Brücken, S. Schuster, & M. Wienberg (Eds.), *Aspects of (Post)Colonial Linguistics: Current
 1094 Perspectives and New Approaches*. Berlin, Boston: De Gruyter. 27–52.
- 1095 Sumbatova, Nina R. & Lander, Yury A. 2014. Darginskij govor selenija Tanty: grammatičeskij očerk,
 1096 voprosy sintaksisa [The Dargwa dialect of Tanty: A grammatical sketch, syntactic issues]. Moscow:
 1097 Jazyki slavjanskix kul'tur.
- 1098 Sussex, Ronald. (1993). Slavonic languages in emigration. In B. Comrie & G. G. Corbett (Eds.), *The
 1099 Slavonic Languages*. London: Routledge. 999–1035.
- 1100 Švedova, Natalija Ju. (1980). *Russkaja Grammatika. Tom II* [Russian Grammar. Volume II]. Moscow:
 1101 Nauka.
- 1102 Szmrecsanyi, Benedikt. (2013). The great regression. Genitive variability in Late Modern English news
 1103 texts. In K. Börjars, D. Denison, & A. K. Scott (Eds.), *Morphosyntactic Categories and the
 1104 Expression of Possession*. Amsterdam/Philadelphia: John Benjamins. 59–88.
- 1105 Szmrecsanyi, Benedikt, Biber, Douglas, Egbert, Jesse, & Franco, Karlien. (2016). Toward more
 1106 accountability: Modeling ternary genitive variation in Late Modern English. *Language Variation
 1107 and Change* 28(1):1–29.
- 1108 Tagliamonte, Sali A. & Baayen, Harald R. (2012). Models, forests, and trees of York English: Was/were
 1109 variation as a case study for statistical practice. *Language Variation and Change* 24(2):135–78.
- 1110 Testelec, Yakov G. (1998). Word order in Daghestanian languages. In A. Siewierska (Ed.), *Constituent
 1111 Order in the Languages of Europe*. Berlin/New York: Mouton de Gruyter. 257–80.
- 1112 Thomason, Sarah G. (2001). *Language Contact*. Edinburgh: EUP.
- 1113 ———. (2009). Why universals versus contact-induced change. In M. Filppula, J. Klemola, & H.
 1114 Paulasto (Eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of
 1115 English and Beyond*. London: Routledge. 349–64.
- 1116 Thomason, Sarah G. & Kaufman, Terrence. (1988). *Language Contact, Creolization, and Genetic
 1117 Linguistics*. Berkeley (CA): University of California Press.
- 1118 Viti, Carlotta. (2008). Genitive word order in Ancient Greek: A functional analysis of word order
 1119 freedom in the noun phrase. *Glossa* 84:203–38.
- 1120 Wasow, Thomas. (1997). Remarks on grammatical weight. *Language Variation and Change* 9(1):81–
 1121 105.
- 1122 Zemskaja, Elena A. (1987). *Russkaja razgovornaja reč': lingvističeskij analiz i problemy obučenija*
 1123 [Colloquial Russian: linguistic analysis and learning problems]. Moscow: Russkij jazyk.
- 1124
- 1125