EDITED BY

MARIA
POLINSKY

The Oxford Handbook *of*
LANGUAGES
OF THE CAUCASUS

THE OXFORD HANDBOOK OF

LANGUAGES OF THE CAUCASUS

THE OXFORD HANDBOOK OF

LANGUAGES

OF THE

CAUCASUS

Edited by

MARIA POLINSKY

OXFORD

UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

In memory of Helma van den Berg (1965–2003): The light will not go out…

Table of Contents

Editor

·····················

Maria Polinsky is Professor of Linguistics and Associate Director of the Language Science Center at the University of Maryland. She has done extensive primary work on several language families, in particular, on languages of the Caucasus: Nakh-Dagestanian, Norwest Caucasian, and Kartvelian. Her research emphasizes the importance of lesser-studied languages for theoretical linguistics.

Abbreviations

:	stem marker
#	number
1, 2, 3	first, second, third person
I, II, III, etc.	gender categories (Nakh-Dagestanian family)
A/B/C	Set A/B/C agreement affixes (Kartvelian family)
A	agent-like argument of a two-place verb
ABL	ablative
ABS	absolutive
ABST	abstract
ACC	accusative
ACT	active
ACTL	actualizing
AD	near, by (reference point/localization)
ADD	additive
ADDR	addressee-oriented
ADJ	adjective
ADJZ	adjectivizer
ADV	adverb, adverbial
AFF	affective case
AFFIRM	affirmative
AGR	agreement
AGT	agent
AL	alienable
ALL	allative
alv.-pal.	alveolo-palatal
alveo.	alveolar
AN	action nominal
ANIM	animate
ANT	anterior
ANTE	in front (reference point/localization)
AOR	aorist
AP	antipassive
APPL	applicative
APPR	approximative

APUD	near, by, at (reference point/localization)
ART	article
ASSOC	associative
AT	at (reference point/localization)
ATTR	attributive
AUG	augment
AUX	auxiliary
B	B gender (Chechen and Ingush)
BEN	benefactive
bilab.	bilabial
CAUS	causative
CF	counterfactual
CHEZ	adverb formative meaning 'at the home of…', 'at…'s place' in Ingush (Fr. *chez*)
CIRC	circumferential
CISL	cislocative
CL	classifier
CMPL	completive
COLL	collective
COMIT	comitative
COMP	complementizer
COMPR	comparative
CON	conative
Con	constraint inventory
CONC	concessive
COND	conditional
CONJ	conjunctive
CONT	on a vertical surface/in an attachment configuration (reference point/ localization)
CONTR	contrastive
COORD	coordination
COP	copula
CV	characteristic vowel
CVB	converb
D	D gender (Chechen and Ingush)
DAT	dative
DEB	debitive
DECL	declarative
DEF	definite
DEIC	deictic
DEM	demonstrative
DEM.ADDR	demonstrative 'close to the addressee'

DEM.DOWN	demonstrative 'lower than the speaker's reference point'
DEM.NEAR	demonstrative 'near the speaker'
DEM.SP	demonstrative 'close to the speaker'
DEM.UP	demonstrative 'higher than the speaker's reference point'
DET	determiner
df	degrees of freedom
DIR	directional, directive
DIST	distal
DISTR	distributive
DO	direct object
DOM	differential object marking
DUBIT	dubitative
DUR	durative
DYN	dynamic
EANC	Eastern Armenian National Corpus
EC	euphonic consonant
EGO	egophoric
ELAT	elative
EM	extension marker
EMPH	emphatic
EP	epenthetic
EQ	equative
ERG	ergative
ESS	essive
EV	euphonic vowel
EVID	evidential
EVT	eventual
EXC	excessive
EXCL	exclusive
EXP	experiencer
EXST	existential
EZF	ezafe
F	feminine
FACT	factive
FCL	facilitive
FIN	finalis
FOC	focus
FUT	future
FV	final voicing
GEN	genitive
GER	gerund
glot.	glottal

GM	gender marker
GNT	general tense
H	human
HAB	habitual
HBL	habilitative
HG	Harmonic Grammar
HOR	horizon of interest
HORT	hortative
IAM	iamitive
IDEOPH	ideophone
IE	Indo-European
IMP	imperative
IMPERS	impersonal
IMPRF	imperfect
IN	inside hollow space (reference point/localization)
INACT	inactive
INAL	inalienable
INANIM	inanimate
INC	inceptive
INCH	inchoative
INCL	inclusive
IND	indicative
INDEF	indefinite
INDIR	indirect
INF	infinitive
INFER	inferential
INS	instrument(al)
INTENS	intensive; intensifier
INTER	inside a mass/solid substance, amorphous space (reference point/localization)
INTJ	interjection
INVOL	involuntative
IO	indirect object
IPFV	imperfective
IRR	irrealis
ITER	iterative
ITF	intentional future
ITR	intransitive
J	J gender (Chechen and Ingush)
K	Kartvelian
L	local (1st or 2nd) person
l.-d.	labio-dental

labial.	labialized
LAT	lative
lat. appr.	lateral approximant
lat. fric.	lateral fricative
lat. lab.	lateral labialized
LIKE	similarity
LNK	linker
LOC	locative
LOG	logophoric
LS	lexical stem (within complex verbs)
LV	light verb
M	masculine
MAL	malefactive
MDT	meditative
MED	medial demonstrative
MIR	mirative
MNR	manner
MOD	modal
ms	millisecond
MSD	masdar
N	neuter
n	non-
NARR	narrative
ND	Nakh-Dagestanian
NDEB	Non-Derived Environment Blocking
NEG	negation, negative
NFC	non-finite conditional
NMLZ	nominalizer, nominalization
NOM	nominative
NONDUM	'not yet' marker
NUM	number, numerative
NWC	Northwest Caucasian
OBJ	object
OBL	oblique
OCP	Obligatory Contour Principle
OPT	optative
ORD	ordinal
OS	oblique stem
OT	Optimality Theory
P	patient-like argument of a two-place verb
pal.	palatal
palat.	palatalized

PASS	passive
PCC	Person-Case Constraint
PFV	perfective
phar.	pharyngeal
phar./epigl.	pharyngeal/epiglottal
phar.&lab.	pharyngealized and labialized
pharyng.	pharyngealized
PL	plural
PLUPRF	pluperfect
PN	person-number ending
POSS	possessive
POST	behind (reference point/localization)
post-al.	post-alveolar
POSTP	postposition
POT	potential
PQ	polar question
PR	possessor series of personal prefixes
PREF	prefix
PRET	preterite
PRF	perfect
PRL	prolative
PROG	progressive
PROH	prohibitive
PROX	proximal
PRP	preposition
PRS	present
PRVT	privative
PST	past
PTCL	particle
PTCP	participle
PURP	purposive
PV	preverb
Q	question
QNT	quantifier
QUOT	quotative
R	root
R.EXT	root extension
RDP	reduplication
RE	refactive
REC	reciprocal
REF	referential

REFL	reflexive
REL	relative; relativizer; relative pronoun, relative pronoun affix
REP	reportative
REPET	repetitive
RES	resultative
rev	reversative
RQ	rhetorical question
RT	pre-root vowel
S	single argument of a one-place verb
SAP	speech act participant
SBJV	subjunctive
SBST	substantivizer
SEP	separation (spatial)
SG	singular
SIM	simultaneous
SM	series marker
SML	similative
SPRESS	superessive
SSP	Sonority Sequencing Principle
ST	stative
SUB	under (reference point/localization)
SUBJ	subject
SUBORD	subordinator
SUBST	substitutive
SUFF	suffix
SUPER	above, over, on top (reference point/localization)
TAG	tag question
TAM	tense, aspect, mood
TAME	tense, aspect, mood, evidentiality
TEMP	temporal
TH	thematic element
TR	transitive
TRAL	translocative
TRANS	translative
TS	thematic suffix
uvul.	uvular
V	V gender (Chechen and Ingush)
vel.	velar
VERIF	verificative
VERS	versionizer
VN	verbal noun

VOC	vocative
VOT	voice onset time
WH	*wh*-agreement
WHQ	*wh*-question
WIT	witnessed

Notes on Contributors

Peter Arkadiev holds a PhD in theoretical, typological, and comparative linguistics from the Russian State University for the Humanities and a Habilitation degree from the Russian Academy of Sciences. Currently he is a Senior Researcher at the Institute of Slavic Studies of the Russian Academy of Sciences and an Assistant Professor at the Russian State University for the Humanities. His fields of interest include language typology and areal linguistics, morphology, case and alignment systems, tense-aspect, Baltic languages, and Northwest Caucasian languages. He co-edited "Contemporary Approaches to Baltic Linguistics" (with Axel Holvoet and Björn Wiemer), "Borrowed Morphology" (with Francesco Gardani and Nino Amiridze), both published by De Gruyter Mouton in 2015, and "The Complexities of Morphology" (with Francesco Gardani), published by the Oxford University Press in 2020.

Gašper Beguš is an Assistant Professor of Linguistics University of California, Berkeley. His work focuses on experimental and computational phonology, phonetics, and historical linguistics. His PhD from Harvard University (2018) proposes a model that combines experimental phonology with statistical modeling of sound change to better understand cognitive and historical influences on phonological grammar.

Oleg Belyaev is a Lecturer in the Department of Theoretical and Applied Linguistics at Moscow State University, a Research Fellow at the Department of Linguistic Typology at the Institute for Linguistics of the Russian Academy of Sciences, and a Research Fellow at the Pushkin Russian Language Institute in Moscow. His research interests include languages of the Caucasus (especially Ossetic and Dargwa), the syntax and semantics of clause linkage, and the typology and theory of morphological case systems. He is one of the editors of the Languages of the Caucasus series at Language Science Press (http://langsci-press.org/catalog/series/loc).

Natalia Bogomolova is a PhD candidate in the Department of Linguistics at the University of Bamberg and a Research Fellow in the Department of Caucasian Languages at the Institute for Linguistics of the Russian Academy of Sciences in Moscow. She holds a PhD in Russian Literature from Moscow State University. Her work concentrates on description and theoretical analysis of Tabasaran and other Lezgic languages. She is currently interested in the syntax of ergativity, clitic doubling, binding, and indexical shift.

Lena Borise is a Research Fellow at the Research Institute for Linguistics (Budapest, Hungary). Her work focuses on the acoustic manifestation of prosodic prominence,

syntactic realization of information structure, and the interaction between the two. In her PhD dissertation at Harvard University (Borise, 2019b), she investigated the prosodic and syntactic properties of focus in Georgian.

Marina Chumakina is a Research Fellow in the Surrey Morphology Group at the University of Surrey. Her work focuses on morphology and syntax of Nakh-Dagestanian languages, especially agreement. She has done extensive fieldwork on the Archi language resulting in an online Archi dictionary (Chumakina, Brown, Corbett, & Quilliam, 2007a) and a collection on Archi agreement (Bond, Corbett, Chumakina, & Brown, 2016).

Michael Daniel is a Professor at the School of Linguistics and a Research Fellow in the Linguistic Convergence Laboratory, both at the Higher School of Economics (HSE) in Moscow. He received his PhD from the Russian State University for the Humanities in 2001. His interests include fieldwork in Dagestan and typology of nominal categories.

Ömer Demirok is an Assistant Professor in the Department of Linguistics at Boğaziçi University, Istanbul. He received his PhD in Linguistics from the Massachusetts Institute of Technology in 2019. His primary research interests include syntax, formal semantics, and their interface. He has done fieldwork on the Pazar (At'ina) dialect of Laz and on heritage Georgian varieties spoken in Turkey.

Nina Dobrushina is a Professor at the School of Linguistics and Head of the Linguistic Convergence Laboratory at the Higher School of Economics (HSE) in Moscow. She graduated from Moscow State University and received her PhD from the Russian State University for the Humanities in 1995. She works on languages of Dagestan, multilingualism, typology of mood and modality, and language variation. In 2016, she published a monograph on the Russian subjunctive. She has done fieldwork on Archi, Mehweb, Rutul, and some other languages.

Ömer Eren is a PhD student in the Linguistics Department at the University of Chicago. He holds an MA from Boğaziçi University in Turkey (2017). The primary focus of his work is on morphology, syntax, and their interface in Turkic and Caucasian languages. He is mainly interested in the structure of nominals and spatial constructions.

David Erschler is a Lecturer in the Department of Foreign Literatures & Linguistics at Ben-Gurion University of the Negev. He received his PhD in Mathematics from Tel Aviv University in 2007, and a PhD in Linguistics from the University of Massachusetts at Amherst in 2018. His research interests include syntactic and morphological typology, ellipsis, fieldwork, and diachronic syntax. He has done fieldwork on Ossetic, Georgian, and Svan.

Steven Foley is a postdoctoral researcher at Princeton University. With interests in both formal syntax and psycholinguistics, he has conducted theoretical and experimental research on Georgian and on Zapotec languages. This work has touched on issues including agreement, clitic pronoun movement, and relative-clause processing.

Diana Forker is a Full Professor of the Caucasus Studies at the University of Jena. Her work focuses on Caucasian languages from a functional-typological perspective, in particular morphosyntax, language contact, and sociolinguistics. She has recently published "A Grammar of Sanzhi Dargwa" (Language Science Press). Together with Oleg Belyaev and a team of experts in Caucasian languages, she is building an electronic lexical database of Caucasian languages (LexCauc).

Victor A. Friedman is Andrew W. Mellon Distinguished Service Professor Emeritus in the Humanities at the University of Chicago and Honorary Adjunct at La Trobe University. He holds a PhD in both Linguistics and in Slavic Languages and Literatures from the University of Chicago (1975). He has received the Annual Award for Outstanding Contributions to Scholarship from both the American Association of Teachers of Slavic and East European Languages (2009) and the Association for Slavic, East European, and Eurasian Studies (2014). His main research interests are all linguistic aspects relating to languages of the Balkans and the Caucasus.

Dmitry Ganenkov is a Research Fellow in the Department of English and American Studies at the Humboldt University of Berlin and a Senior Research Fellow at the Laboratory for Languages of the Caucasus at the Higher School of Economics (HSE) in Moscow. He received his PhD in Linguistics in 2005 from Moscow State University. His work focuses on documentation, description, and syntactic analysis of Nakh-Dagestanian languages. His theoretical interests include ergativity, agreement, binding, and obligatory control.

Konstantin Kazenin has a PhD in Linguistics from Moscow State University. He worked at the Department of Theoretical and Applied Linguistics of Moscow State University from 1997 to 2012, researching languages of Dagestan (Lak, Tsakhur, Bagwalal) and syntactic typology. In 2012, he joined the Russian Academy for National Economy and Public Administration, where he is the head of the Department for Regional Studies. His current research concentrates on demography (mainly fertility dynamics) in the Caucasus and Central Asia, as well as labor migration from Central Asia to Russia. He also serves as Associate Professor in the Department of Demography of the Higher School of Economics (HSE) in Moscow. He is a member of the European Association for Population Studies.

Erwin R. Komen received his PhD at the Radboud University Nijmegen, and combines his job writing software programs for humanities researchers with his work as a linguistics consultant at SIL International. His recent web application "Cesar" facilitates syntactic research in, among others, the Nijmegen parsed corpus of modern Chechen.

Yuri Koryakov is a Senior Researcher in the Department of Areal Linguistics in the Institute for Linguistics of the Russian Academy of Sciences in Moscow. His research interests include Caucasian sociolinguistics and writing systems, and, most importantly, language cartography and taxonomy. He is author or co-author of several monographs, including the "Atlas of Caucasian Languages" (2002).

Yury Lander is an Associate Professor in the School of Linguistics at the Higher School of Economics (HSE) in Moscow. He has worked on both Nakh-Dagestanian and Northwest Caucasian languages, including Udi, Dargwa, Adyghe (West Circassian), Kabardian (East Circassian), and Abaza, as well as on some Austronesian languages. He is particularly interested in various topics in syntactic and morphological typology, including relativization, possessive constructions, and polysynthesis.

Timur Maisak is a Senior Research Fellow at the Institute for Linguistics of the Russian Academy of Sciences and a Research Fellow at the Linguistic Convergence Laboratory of the Higher School of Economics (HSE) in Moscow. He obtained his PhD in Linguistics from Moscow State University in 2002. He has conducted research on Nakh-Dagestanian languages since the 1990s, mainly on the Lezgic and Andic branches of the family. His research interests include language documentation and description, typology of verbal categories, and grammaticalization theory.

Zarina Molochieva is a Lecturer at the Department for General Linguistics at the University of Kiel (Germany). She received her PhD from the University of Leipzig in 2011. Her research interests lie in verbal morphology, especially tense, aspect, and mood systems, as well as language documentation and description.

Johanna Nichols is Professor Emerita at the University of California, Berkeley. She received her PhD from the University of California, Berkeley, and has taught there for most of her career. She works on typology, historical linguistics, languages of the Caucasus (especially Ingush), and language spreads in northern Eurasia.

Brian O'Herin has conducted research on the morphology and syntax of Abaza since the 1980s, receiving his PhD in Linguistics from the University of California at Santa Cruz in 1995. He taught at Biola University before serving as the International Linguistics Coordinator for SIL International, and has recently returned to field research, focusing on syntax and text linguistics.

Balkız Öztürk is an Associate Professor of Linguistics at Boğaziçi University, Istanbul. She received her PhD from Harvard University in 2004. Her research interests include the interface between syntax, morphology, and the lexicon. She focuses on Altaic and South Caucasian languages. She is the author of the monograph "Case, Referentiality and Phrase Structure" and has co-edited the volumes "Exploring the Turkish Linguistic Landscape", "Morphological Complexity Within and Across Boundaries", and "Pazar Laz".

Maria Polinsky is a Professor of Linguistics and Associate Director of the Language Science Center at the University of Maryland, College Park. Her main interests are in theoretical syntax, with an emphasis on cross-linguistic variation. She is also interested in the integration of experimental methodologies in linguistic research. She has done extensive primary work across several language families, including languages of the Caucasus.

Alexander Rostovtsev-Popiel is a Research Fellow at the Institute for Slavic, Turkic and Circum-Baltic Studies of the Mainz University. He received his PhD in Linguistics, for a dissertation on grammaticalization in Kartvelian, at Frankfurt University in 2012. From 2017 to 2019, he was a Postdoctoral Researcher at Collège de France. He specializes in morphology, morphosyntax, and pragmatics of Kartvelian languages.

Nina Sumbatova is a Professor of Linguistics at the Russian State University for the Humanities in Moscow. She received her advanced degrees from the Institute for Linguistics at the Russian Academy of Sciences. Her research interests lie in the morphology and syntax of Dargwa and other Nakh-Dagestanian languages, as well as theoretical syntax and linguistic typology. She is the author of two descriptions of Dargic languages: "Itsari Dargwa" (2003, with Rasul Mutalov) and "Tanti Dargwa" (2014, with Yury Lander).

Yakov G. Testelets (Testelec) is a Professor of Linguistics at the Russian State University for the Humanities and head of the Department of Caucasian Languages in the Institute for Linguistics at the Russian Academy of Sciences. His work has focused on the grammar of languages of the Caucasus, especially Svan (Kartvelian), Adyghe (West Caucasian), Tsakhur, Bezhta, and Avar (Nakh-Dagestanian), problems of grammatical theory and typology (anaphora, linear order, grammatical asymmetries, ergativity, case), and the syntax of Russian.

MAP 1. Caucasus: Administrative division

MAP 2. Language families of the Caucasus

Languages of Caucasus

© Koryakov Yu.B., 2020

Adyghe · Assyrian · Russian · Kalmyk

Abaza · Nogay · RUSSIA · Nogay · Avar, Russian, Dargwa, Kumyk, Lak, etc.

Kabardian · Karachay-Balkar · Kabardian · Chechen · CASPIAN SEA

Adyghe · ABKHAZIA · Svan · Ingush · Kumyk

Hamshen Armenian · Ossetic · ANDIC · Avar

Abkhaz · Megrelian · SOUTH OSSETIA · TSEZIC · DARGIC · Lak

BLACK SEA · GEORGIA · Georgian · B · Azerbaijani · LEZGIC

Hamshen Armenian · TBILISI · Khinalug

Laz · Turkish · ARMENIA · Udi · Tat · AZERBAIJAN

TURKEY · Kurdish · YEREVAN · Azerbaijani

Armenian · ARTSAKH · Azerbaijani

Kurdish · Azerbaijani · Azeri · IRAN · Talysh · BAKU

ABKHAZ-ADYGHE	KARTVELIAN
Adyghe, Kabardian	Mingrelian, Laz
Abkhaz, Abaza	Georgian
	Svan
NAKH-DAGESTANIAN	INDO-EUROPEAN
Chechen, Ingush	Russian
B Bats	Armenian
Avar	Iranian: Ossetic, Kurdish, Tat, Talysh
ANDIC	
TSEZIC	TURKIC: Nogay, Karachay-Balkar,
Lak	Nogay, Kumyk, Azeri, Turkish
DARGIC	MONGOLIAN: Kalmyk
LEZGIC	SEMITIC: Neo-Aramaic
Khinalug	

MAP 3. Languages of the Caucasus

Nakh-Dagestanian languages

© Koryakov Yu.B., 2020

Chechen

Andi

DAGESTAN

Kumyk

2
1
Karata
Khunzakh

Chamalal

3
4

Tindi

5

5

Khwarshi

10

Tsez

7 Hunzib

Bezhta

6

North
Dargwa

Tsudaqar

D A R

Kumyk

Avar

12

Lak

13
14

SW. Dargwa

G Muira

W

15
16

Kaitag

17

18

A

Azerbaijani

Derbent

CASPIAN
SEA

Archi

8   9

11

15

Chirag

Lak

Agul

Tabasaran

Avar

Tsakhur

4

Zaqatala

Rutul

Rutul

Lezgi

Azerbaijani

Lezgi

Quba

GEORGIA

Şəki

Kryz

Khinalug

Budukh

Azerbaijani

ARMENIA   AZERBAIJAN

1 Godoberi
2 Botlikh
3 Bagvalal
4 North Akhvakh
5 South Akhvakh
6 Batsbi
7 Hinukh
8 Bezhta
9 Hunzib
10 Sagada
11 Udi
12 Mehweb
13 Hapshima
14 Usisha-Butri
15 Kubachi-Ashti
16 Amuzgi-Shiri
17 Sanzhi-Itsari
18 Shari

MAP 4. Nakh-Dagestanian languages

INTRODUCTION

MARIA POLINSKY

I.1. NAVIGATING THE AREA

THE Caucasus is a relatively small land mass between two seas: the Black Sea on the west and the Caspian Sea on the east. Its northernmost area includes the Great Caucasus mountain range, and its southernmost part shares a border with Turkey and Iran. The Caucasus is separated from Russia by the Kuban and Terek Rivers in the north and is bound by the Kura and Araxes Rivers in the south. Famous for its dizzying cultural and linguistic diversity, this small, rectangular region of mountains (including Mount Elbrus and Mount Kazbek, which are best-known), hills, plateaus, valleys, and meadows has long been the homeland to many ethnic groups. "The ethnic complexity of the Caucasus is unequalled in Eurasia, with nearly sixty distinct peoples, including Russians and Ukrainians" (Colarusso, 2009). Rarely does an overview fail to mention the nickname given to the Caucasus by medieval Arab historians, "a mountain of tongues" (see Catford, 1977; Chumakina, 2011a, among others).

Traditionally the Caucasus is divided into two main parts: the North Caucasus (Ciscaucasus, Ciscaucasia) and the South Caucasus (Transcaucasus, Transcaucasia). While about a hundred or so languages are spoken in the Caucasus, there are three major language families that exist solely in the Caucasus and do not have any member languages outside the area (various late diasporas do not count here). These three families are considered indigenous. Sometimes, the phrase "languages of the Caucasus" or, more accurately, "Caucasian languages" refers to these languages only.[1] Two of these

---

[1] See Comrie (2005) for the terminological distinction between "languages of the Caucasus" and "Caucasian languages," and see also chapter 1. The indigenous status of Caucasian languages does not prevent speakers of individual languages of these families from arguing with each other about who got there first. This is a difficult topic, associated with many political and cultural issues, often confounded by a lack of clear historical data. Since this *Handbook* focuses on the linguistic richness of the area in modern times, it does not include any discussion of territorial origins or genetics. Genetic investigations addressing the migration history in the area have appeared in the last decade (Balanovsky et al., 2011;

indigenous families are found in the North Caucasus; the third is in the south. The north can be conveniently divided into the northwest, home of the Northwest Caucasian (Abkhaz-Adyghe) family, and the northeast, home of the Nakh-Dagestanian family.[2] The south is where languages of the Kartvelian (South Caucasian) family are spoken. Both the Northwest Caucasian family and the Kartvelian family are small in terms of member languages. The former consists of Abkhaz, Abaza, Kabardian and Adyghe (these two are often combined under the umbrella term "Circassian"), and Ubykh. The Kartvelian family includes Georgian, Megrelian, Laz, and Svan. On the other hand, the Nakh-Dagestanian family includes many more languages. As its name suggests, this family is comprised of two main branches: Nakh and Dagestanian. While the Nakh languages form a single genealogical grouping,[3] the languages traditionally called Dagestanian do not—this term reflects common geography rather than early branching in the history of the family.[4]

Researchers looking for long-range linguistic comparisons place Kartvelian languages in the Nostratic family (Bomhard, 2008; Illich-Svitych, 1971, among others) and connect the Northwest Caucasian and Nakh-Dagestanian families to Sino-Tibetan (Nikolaev & Starostin, 1994). No matter how we look at it, the three indigenous language families do not form a genealogical unit.[5] Why, then, treat them together? Bernard Comrie offers an explanation, relying on traditional training and common geography: "One reason is historical, namely that the training of specialists has tended to be across the range of Caucasian languages, even if with greater specialization in just one of the three families. This also makes sense practically, for instance in that students of these languages share certain prerequisites, such as at least a reading knowledge of Russian, often also of Georgian. But perhaps more important than this is the fact that these languages occupy a more or less contiguous geographical area at the boundary of Europe and Asia as both geographical and cultural entities, an area that is moreover surrounded by representatives of much larger language families . . ." (Comrie, 2005, p. 1).

In addition to the three indigenous families, the Caucasus is home to several languages that belong to families with wider distribution. Most notable among the Indo-European languages are Armenian and Ossetic, whose speakers have long lived in the area. Northern Kurdish and (Judeo)Tat are fading, with fewer and fewer native speakers left.[6] Of the Turkic family, Azerbaijani, spoken in the south, is the largest. Other Turkic languages include Kumyk, Karachay-Balkar, and Noghay. For several other languages of the area, see chapter 1.

Karafet et al., 2016; Wang et al., 2018), but more work remains to be done. Of resources in English, see King (2008) and Forsyth (2013) for the history of the region and Rayfield (2012) for the history of the South Caucasus, with further references therein.

[2]   Here and below, I will be using the most common names of language families and individual languages. For alternative names (of which there are many), see chapter 1 and appendix I.

[3]   See chapters 3 and 8.   [4]   See chapters 1 and 3, for more discussion.

[5]   See also chapter 1.   [6]   See chapter 13.

The maps included with this *Handbook* show the main administrative divisions in the area, the distribution of the main families, and a more detailed distribution of languages within these families.

In an area as compact and densely populated as the Caucasus, multilingualism is more a norm than an exception, and research on language contact among languages of the area has always been very productive. At some point, researchers were even tempted to propose the concept of the Caucasian *Sprachbund* (Chirikba, 2008b; Klimov, 1978; Klimov & Alekseev, 1980; but see Tuite, 1999, for arguments against this approach). The main trends in multilingualism and contact in the Caucasus are discussed in chapter 1, with further references on this topic.

Aside from the many local languages in contact, several other languages have been present in the region, too—by virtue of geography and politics. Located at the peripheries of Turkey, Iran, and Russia, and literally at the crossroads of Europe and Asia, the Caucasus has long been an arena for expansionism and political, military, religious, and cultural rivalries. Until the end of the 18th century, the area was first aligned, politically and culturally, with the Arab world, and later with the Persian and Ottoman Empires. The languages associated with these outside forces left a strong mark within the Caucasus, to the point that numerous Arabic, Turkic, and Iranian (Iranic) borrowings remain throughout the languages of the region.[7] Many words of Middle Eastern origin show up in all of these languages, and it is not always easy to determine if a given loanword comes directly from Arabic, Turkish (or other Turkic languages), Persian, or another Iranian language or traveled from one of these outsider languages to another and then later, to a particular Caucasian language.

The literature on loanwords from Arabic, Turkish, and Iranian languages in Kartvelian languages is quite substantial (Fähnrich, 2007; Gippert, 1990; Klimov, 1998a, and references therein). For loans from Northwest Caucasian into Kartvelian, see Chirikba (1998, 2006) and references therein, and for Nakh-Dagestanian loans in Kartvelian, see Fähnrich (1988, 2007). Studies of Arabic, Turkic, and Iranian loanwords in languages of the North Caucasus are also popular in the local philological tradition. For monographic descriptions of such borrowings into Nakh-Dagestanian languages, see Dzhidalaev (1990), Selimov (2010), Zabitov (2001), and Zabitov and Èfendiev (2001)—these studies include many further references.

Yet another outside language has maintained a formidable and vigorous presence in the region since the 19th century: Russian. In the beginning of the 19th century, the Caucasus was annexed by the Russian Empire (see Baddeley, 1908; Potto, 1887–1889, for the history of the Russian invasion and subsequent annexation). The Russian conquest of the Caucasus was not unlike the settlement of New Zealand by the British or the conquest of the Sahara by the French. The remote, strange, and, at times, bleak landscape seemed squalid and uninhabitable; both its climate and its horticulture were entirely

---

[7] Loans from Turkish dominate Turkic borrowings. Among Iranian borrowings, Persian loans are most noticeable. Throughout this *Handbook*, references to Turkic/Turkish and Persian/Iranian can be found interchangeably.

foreign. The steep mountains did not appeal to the Russian peasant farmers, who were more interested in the rich fields and forests of Siberia. Promises of natural resources and salt mines were played up by the locals, but those remained unfulfilled. And in 1801, oil drilling was not a lucrative undertaking. Instead, this alien terrain attracted vagabonds, criminals, and romantic literati who marveled at the exotic locale. The rest of the Russian settlers were moved forcibly, often as part of army divisions.

Despite reservations, the Russian Empire was drawn to the Caucasus for two reasons. First, the tsars were trying to establish a reliable border with Iran and Turkey, one that they could hold steady. In this regard, the South Caucasus was the real prize, whereas the North Caucasus was viewed as more of a nuisance—the price that had to be paid in order to create a Russian presence at the Iranian and Turkish borders. Second, as a strong Christian nation which considered itself a direct descendant of Byzantium, the Russian Empire sought to protect Christians in the Caucasus, such as Georgians and Armenians (and the less numerous Greeks). For their part, the Georgians and Armenians in the South Caucasus were also looking to align themselves with the Russians for religious reasons, as they were worried that an alliance with the Persians or the Ottoman Empire would force them into Islam. With a heavy heart, the Georgian Bagrationi dynasty accepted the inclusion of their lands in the Russian empire as the lesser of two evils.[8]

The time that has passed since Russia's conquest of the Caucasus has not been easy. Following the disintegration of the Soviet Union in 1991, periods of independence have been punctuated by vicious military fighting—such as a series of brutal Chechen wars (see German, 2003, and references therein; see also chapter 2) and the Russo-Georgian war of 2008. Political and military turmoil aside, the linguistic presence of Russian has remained significant throughout the area since the 19th century, especially in the North Caucasus where Russian has displaced a dozen or so local languages that used to be *linguae francae,* becoming the main common language (see chapter 1). Russian is "considered by many not to be a truly 'foreign' language (like French, German or English), but rather a sort of second native language (regardless of how well they actually spoke it)" (Blauvelt, 2013, p. 3).

The role played by Russian is evident from the local migration patterns. As soon as speakers of a local language move to a more urban setting (which is often associated with migration from the highlands to the multiethnic lowlands), Russian becomes dominant. This ongoing switch to Russian has consequences both for Russian and for local languages. First, as Russian remains a prestigious, important language in the area, one associated with upward mobility, local varieties of Russian emerge (Belikov, 2011; Daniel, Dobrushina, & Knyazev, 2010; chapter 1 of this volume). In the Soviet days, such varieties of Russian were mostly ignored and considered substandard; current work on these varieties is in its early stages, and they need to be investigated more.

---

[8] Although the Orthodox Christianity shared by the Georgians and Russians was important in the dialogue between the two nations, Georgian kings also pursued the option of aligning with the Catholic Church (Lang, 1957).

Second, despite the fact that many censuses indicate large numbers of speakers for certain languages (see chapters 1 and 2 in this *Handbook*), a significant proportion is represented by semi-speakers or heritage speakers: recessive bilinguals who are more dominant in Russian. Furthermore, quite a few groups in the Caucasus identify themselves based on ethnicity and may state that they speak the corresponding language, when really, they only know a few words (see chapters 1 and 2).[9] The growing dominance of Russian underscores the urgency of studying the languages of the northern Caucasus; the often-misleading numbers of speakers of a given language may give researchers the sense of false comfort concerning linguistic vitality.

Though Russian has supplanted several local languages that used to be widely spoken, at least two languages, Georgian and Armenian, have withstood its pressure. Their endurance in the Russian Empire, and later in the Soviet Union, can be explained in part by the long-standing literary traditions in both languages, not to mention the sheer number of speakers for each. Both the Armenian and Georgian scripts go back to the 5th century (their origins are a point of contention), and medieval chronicles in both languages date back to the 9th century. There is a tremendous body of literature in both languages, which forms a common cultural background for the populations, who have an extremely high literacy rate. In the Soviet Empire, the constitutions of the local republics provided for the use of the titular (local) language and Russian, although Russian was tacitly assumed to be the more important, more prestigious language (Blauvelt, 2013; Slezkine, 1994). The Soviet "ethnophilia" of the 1920s, in which all minority languages and ethnicities were supported, yielded to the policies of the mid-1930s, which supported larger nationalities, especially ones that had titular republics within the Soviet Union. Georgian and Armenian benefited significantly in both periods, becoming the languages of state bureaucracy (Blauvelt, 2014).

Around the mid-1930s, the central Soviet government decided that Georgia and Armenia would serve as the model "advanced republics" of the Union. As a result, their languages, cultures, and what was called "ethnogenesis" became the focus of all republican academic institutions created by the party and state—including unions of writers, institutes of history, ethnography, literature, archaeology, and so on. This special status played out in many ways. One example can be traced back to the late 1930s, when Georgian and Armenian were able to retain their traditional scripts (granted, they had had these traditional scripts for centuries, as mentioned above). Republican languages that did not have traditional writing systems, but rather, Latin-based orthographies developed in the 1920s, were all required to use the Cyrillic script in the late 1930s (the Azerbaijanis switched back to the Latin orthography in the 1990s, after the fall of the Soviet Union). At the same time, the languages of the minority groups in Georgia and Armenia (Abkhaz and Ossetic in Georgia, Kurdish in Armenia) switched to Georgian and Armenian scripts, respectively.[10]

---

[9] While this tendency is often noted, the actual numbers of semi-speakers or non-speakers who self-identify with a given group are not known.

[10] Ossetic is particularly telling in that regard: in North Ossetia, the writing system was switched to Cyrillic, and in South Ossetia, to a Kartvelian script. (See chapter 1 for a more general discussion of the writing systems used in the area.)

Georgians were unique in openly protesting against the spread of Russian as the Soviet government attempted to change the constitutional status of languages in Georgia, particularly in 1978. The protestors even disregarded the Soviet regime's oppressive policies on demonstrations (Cornell, 2001). Thus, despite the strong Russification of the Soviet empire in the last several decades, state support for titular languages and institutions continued, creating a kind of paradox wherein official scholarly institutes became bastions of national projects.

Even though their allegiance to their own language was unshakeable, the Georgians did not have second thoughts about the subjugation of more minor Kartvelian languages (Laz, Svan, Megrelian), Abkhaz (spoken in the contested Georgian territory), and Ossetic (spoken in another contested Georgian territory), having even fought off official support for the recognition of peoples they considered to be their own ethnic subgroups (Blauvelt, 2014). Nor did the Armenians worry much about the fate of Neo-Aramaic (Assyrian) or Northern Kurdish (spoken by the Yazidi population) in their country. Russian pushed out minority languages in the North Caucasus, but Georgian and Armenian did the same in their respective domains, too.

I.2. A Linguistic Snapshot
of the Caucasus

Since the languages spoken in the Caucasus are diverse and varied, sweeping generalizations about their design are often superficial and incomplete. All of the region's major language families are known for striking characteristics that receive too much attention, often becoming distorted in the process. Mention Circassian or Kabardian and a likely reaction is that these languages have no vowels—a misinterpretation of the claim that the vowels are fully predictable and, therefore, should not be counted as part of the phonemic inventory (see Catford, 1994, 1997; Kumakhov, 1977, and chapter 15 for a discussion). Languages of Dagestan are best known for their prolific use of case forms (which are, in fact, spatial forms of nouns with incorporated postpositions, see chapter 3; Comrie & Polinsky, 1998) or for their gender oppositions, which are more complex than the usual masculine-feminine distinction.[11] Kartvelian languages are famous for their consonant clusters and complex verb forms, often with different argument alignment depending on the tense, aspect, and presence of additional affixes, such as applicatives, in the verb. This *Handbook* intends to show the genuine complexity and diversity in the Caucasus with the goal of shifting researchers' attention away from the few catchy, Guinness-World-Record-type properties, which are much less exotic than they may seem from the outside.

---

[11] There may be three to eight classes depending on the language. See Corbett (1991), and chapters 3, 8, 20 of this volume.

Undeniably, the Caucasus is a phonetician's paradise. Most indigenous languages of the Caucasus have rich consonant systems with three-way distinctions in the laryngeal features of obstruents that include ejective consonants, as well as a rich inventory of post-uvular articulation, especially in Nakh-Dagestanian. Gašper Beguš (chapter 15) provides a detailed account of the main phonetic and phonological properties that characterize the three major families. As proposed by some researchers, the consistent presence of ejectives may constitute an areal feature (Catford, 1977); beyond the three indigenous families, ejectives are found in Ossetic (see chapters 13 and 14), Neo-Aramaic, as well as in some dialects of Kumyk, Azerbaijani, and Armenian (Chirikba, 2008b, p. 44; Maddieson, 2013). This spread is typically accounted for by the influence from the indigenous languages or the substrate.

I have already mentioned the extensive borrowings from Turkic languages, Iranian languages, and Arabic in languages of the Caucasus. Although borrowings are found in most of the world's languages, the pattern employed by the languages of the Caucasus deserves special mention due to its consistency. Words that relate to politics, religion, some professional names, and even some everyday items are among common borrowings. Furthermore, these words are often so tightly integrated into the lexical systems of the languages that it is hard to identify them as loanwords. The spread of Russian has resulted in a great number of Russian borrowings, as well as the integration of international lexica that arrived via Russian. Borrowings often bear a distinctive phonetic signature, for example, with voiceless stops represented by ejectives in Kartvelian, some Nakh-Dagestanian languages, and Armenian, as in Georgian *p'rop'aganda* 'propaganda', *lep't'op'i* 'laptop', Avar *q'alam* 'pencil', Hinuq *mark'a* 'stamp',[12] Mehweb Dargwa *k'amp'it'* 'candy', and so on. Systematic comparative work on phonetic features of loanwords in the Caucasus is still outstanding.

Most languages of the area are head-final: they have postpositions rather than prepositions, and non-finite clauses are predicate-final.[13] At least one language of the area should be described as having SOV word order and no case marking on noun phrases: Abkhaz (Hewitt, 1979a). The absence of case-marking is typically correlated with verb-medial orders (SVO), and Greenberg's Universal 41 specifically states that, "if in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system" (Greenberg, 1963, p. 75). Thus, Abkhaz is relatively unusual in that regard.[14]

In languages of the area, the word order at the main clause level is usually less rigid, and although verb-initial orders are less common, verb-final and verb-medial orders are typical, as shown in example (1). In quite a few languages, the immediate preverbal position is dedicated to focus constituents; this is a recurrent theme in several descriptive chapters and in Diana Forker's chapter on information structure (chapter 24). A rich

---

[12] In Tsezic languages, borrowings from Russian only show the ejective *k'* (Comrie & Khalilov, 2009b).

[13] But see chapter 13, on prepositions in Indo-European languages of the area.

[14] Combining the features "SOV order" and "no case marking" yields 18 languages of 565 instances of SOV listed by Dryer (2013b) in the *World Atlas of Language Structures*.

postverbal periphery (often referred to as the right periphery) is commonly used for encoding various types of backgrounded or newsworthy information, and in that regard, languages of the Caucasus await comparisons with Hindi-Urdu or Turkish, where the syntax of the right periphery has been investigated (see Kural, 1997; Manetta, 2011, among others). A hallmark of head-final languages are complex predicates, formed from a lexical component and a light verb such as 'be' (for intransitives) and 'do' (for transitives); these are very common throughout the area.

In languages of the North Caucasus, we find a clean distinction between clause-medial (non-finite, converbal forms) co-occurring with the single finite predicate of a complex sentence—consider this long example from Agul (Nakh-Dagestanian), where the only finite predicate is the copular form *x-a-j-e*, itself built on a converb.

(1)   Agul
peʔ   ud-u-na,   mert: aq'-u-na   iži-di,   fajš-u-na,
chicken.ABS tear-PFV-CVB clean   do-PF-CVB   good-ADV   bring-PFV-CVB
ha-te   нüjeg-i-ʕ   ʕix-a-s   bašlamiš aq'-u-guna   kitan
EMPH-DEM.DIST pot-OBL-INTER   INTER-put-INF begin   do-PFV-CVB   cat.ABS
x-a-j-e   me   peʔ-ela-k-as.
become-IPFV-CVB-COP   DEM.PROX   chicken-OBL-SUB.CONT-ELAT
'They pluck the chicken, clean it up really well and bring it over, but when they are ready to put it in the pot, the chicken turns into a cat.'

Is this head-final structural design special to the Caucasus? Probably not. Head-final languages dominate the global linguistic landscape. For instance, all over South Asia, Indo-Aryan and Dravidian languages manifest a similar pattern of head-finality, with participial or converbal clauses dependent on the sole finite predicate. Languages of the Caucasus share non-rigid, head-final properties, including the extended right periphery, with the neighboring Persian and Turkish. It may well be that all of these languages have the most insipid word order and, therefore, areal features should not be held responsible for the apparent uniformity.

All things being equal, one would expect to find the predominance of suffixal morphology in a head-final language. And while suffixation is common across languages of the area, agreement exponents appear before the verbal root in most languages of the three Caucasian families. In Northwest Caucasian and Kartvelian, these exponents index person and number[15]; in Nakh-Dagestanian languages, it is primarily gender and number (see chapter 20). Elements that index person, number, or gender do not have the same categorial status in all the languages of the area. Furthermore, for most languages of the area, whether these elements are morphological prefixes or clitics has yet to be determined. Distinguishing between agreement affixes and clitics is not an easy

---

[15] Abkhaz also has gender agreement, also marked before the verb root (Hewitt, 1979a, pp. 103–125; Shaduri, 2006).

task, but an important one, as this differentiation leads to a better understanding of agreement phenomena in languages of the Caucasus, as well as the order of constituents in the verbal complex, and the nature of ergativity—the feature that I will take up next.

Most languages of the area are ergative and lack passive voice constructions, the latter gap a common, albeit not necessary, corollary to ergativity (see Kazenin, 2001c, for a discussion of this commonly assumed correlation). Ergativity is clearly present in the three indigenous families, yet that superficial parallel is where the similarities end (Catford, 1974; Tuite, 1999; and chapter 18 of this volume). Nakh-Dagestanian languages are consistently ergative, in terms of both their case marking and the agreement with the absolutive in gender (noun class). Their ergativity is purely morphological, it has no syntactic consequences; all types of arguments, regardless of case marking and agreement, can undergo extraction, leaving a gap in the base position.

Ergativity is different in Northwest Caucasian languages. In those languages of the family that have overt case marking, noun phrases are marked for absolutive and ergative, and the ergative coincides with the generalized oblique marker (some researchers argue that it is a single marker). Agreement is with the ergative and with the absolutive, in person and number (gender is present in some but not all languages of the family). The pattern of extraction is different from Nakh-Dagestanian and Kartvelian; in Northwest Caucasian languages, only absolutive arguments can undergo extraction with a gap and no change in the verb form. That characterizes them as syntactically ergative—unlike languages of the other two families.

Finally, in Kartvelian, the ergative appears only in a subset of tense-aspect-mood forms (in Georgian, in the aorist-optative group of TAM forms; see Nash, 2017b, for an analysis). And Kartvelian agreement, famous in its own right for its remarkable complexity, follows the nominative-accusative pattern and tracks only person and number features (see chapter 20).[16] Kartvelian ergativity is thus quite different from the more familiar patterns (of which Nakh-Dagestanian ergativity is probably the textbook case), and some researchers classify Kartvelian languages as having active-inactive rather than ergative case alignment, although the reasons for such an analysis may differ (Harris, 1981; Hewitt, 1987a; Klimov, 1973; and see footnote 16). The main argument for classifying these languages as active-inactive has to do with a large number of verbs that can traditionally be thought of as intransitive ('dance', 'scream', 'yawn') which however have their sole argument marked the same way as a regular transitive subject; in the meantime, the more patient-like arguments of intransitive verbs are marked as transitive objects. This approach, which is more valid for the languages of the family other than Georgian, is reflected in the survey chapter on Kartvelian (chapter 11); but see chapter 18 where these languages are viewed as pretty much middle-of-the-road split-ergative. Clearly the final word on this issue is still to come, and if we want to go beyond just

---

[16] Using more idiosyncratic criteria, Klimov and Alekseev (1980) examine ergativity in all three families and conclude that the Northwest Caucasian languages are the most prototypically ergative, Nakh-Dagestanian languages have elements of nominative-accusative strategies, and Kartvelian languages represent a combination of active, ergative, and nominative types.

naming a particular pattern it is important to operationalize the criteria which define an alignment as ergative-absolutive or active-inactive.

The majority of languages in the Caucasus also have extensive pro-drop. Unlike the better-known pro-drop languages, not only subjects but also direct objects and other non-subject arguments in Caucasian languages can be freely omitted as long as they are recoverable from discourse. It is common to associate pro-drop with rich agreement, and though many languages of the area may have rich agreement (as I mentioned earlier, it is not always clear whether this is agreement or cliticization), pro-drop is also present in languages that lack agreement, such as Lezgian or Agul. Although pro-drop in languages of the Caucasus has been documented (it is hard to miss!), it has not been fully explored yet.

Meanwhile, there are at least two main directions of future research on the nature of pro-drop in languages of the Caucasus. The first one has to do with licensing mechanisms and identification of the null pronominal. Is it due to rich agreement—in other words, are these languages akin to Romance with regard to pro-drop (see Rizzi, 1986)— or are the null pronominals identified by their association with a discourse topic, in a pattern similar to the one claimed for Chinese (see Huang, 1989, 1991)?

The second avenue of research involves patterns of pronominal reference and resolution. Such patterns have been studied in the more familiar Romance languages, where only subjects can be deleted. For Romance, researchers have proposed that null pronouns are preferentially linked to subject antecedents and overt pronouns to antecedents in lower structural positions (Carminati, 2005). Thus, in the Spanish example in (2), the null pronoun in the second clause is preferentially interpreted as referring to the subject, and the overt pronouns *el*, as referring to the object:

(2)   Spanish

    a. *Juan$_i$ pegó   a   Pedro$_k$.   pro$_{i>k}$ está   enfadado.*
       Juan hit   PRP   Pedro   be.PRS.3SG   angry.M

    b. *Juan$_i$ pegó   a   Pedro$_k$.   Él$_{k>i}$ está   enfadado.*
       Juan hit   PRP   Pedro   he   be.PRS.3SG   angry.M

      'Juan hit Pedro. He is angry.' (Keating, Jegerski, & Van Patten, 2016, p. 38)

Since all arguments can be dropped in languages of the Caucasus, what strategies of pronominal reference can we expect? Consider the following example, wherein both the subject and the object are dropped in the second clause, and the clause is ambiguous. So far there has not been any work on strategies of pronominal reference in the Caucasus, and this line of research is promising in that it can bring together issues in theoretical syntax and sentence processing.

(3)   Georgian
    sap'rezident'o   debat'-eb-ši   beridze$_i$-m   gelašvili$_k$
    presidential   debate-PL-LOC   Beridze-ERG   Gelashvili.NOM

uk'mexad   ga-a-k'rit'ik'-a.
harshly   pv-vers-criticize-aor.3sg.3sg
amit'om   pro$_{1SG}$ pro$_{i/k}$ ar   ar a-v-i-rčev.
because.of.that   neg pv-1sg-vers-choose.fut
'At the presidential debates, Beridze$_i$ harshly criticized Gelashvili$_k$.
For that reason, I won't vote for him$_{i/k}$.'

In a number of languages of the area, quantifier phrases are built on uniform indeterminate bases (either full words or stems) that are invariable across different categories, a paradigm that is familiar from Japanese (Haspelmath, 1997; Kuroda, 1965; Nishigauchi, 1990; Shimoyama, 2006). These indeterminate bases combine with an additional morphological exponent (which is typically analyzed as encoding a semantic operator). Depending on the exponent they combine with (including the null one), indeterminate phrases can take on a number of interpretations: interrogative, existential, universal, comparative, negative, negative-polarity, free-choice, and so on. Usually the bare forms have the interrogative interpretation. Consider the following paradigm from Svan (David Erschler, personal communication):

Table I.1 Indeterminate Expressions in Svan

	interrogative	existential	n-words
person	*jær*	*erwa:le*	*dær*
thing	*mæj*	*ma:le/mo:le*	*ma:mgweš/demgwaš*
place	*ime*	*imwa:le*	*deme*
time	*šoma*	*šomwa:le*	*demčik*

Unified or close to unified paradigms of indeterminate expressions are found in most Nakh-Dagestanian languages (see Tatevosov, 2002, for Godoberi, Lak, and Tsaxur; Kibrik, Kazenin, Lyutikova, & Tatevosov, 2001, pp. 165–167 for Bagvalal; Polinsky, 2015b, for Tsez) and in Armenian and Ossetic (Haspelmath, 1997, pp. 281–282). Kartvelian languages have a mostly uniform paradigm for interrogative, existential, negative, and free-choice expressions, but their universal pronouns often have different forms. Northwest Caucasian languages have a partially unified paradigm, with universal and free-choice expressions derived from interrogatives (Nikolaeva, 2012).

Indeterminate expressions raise a number of important questions with respect to quantification, syntactic displacement, or focus, and the addition of Caucasian language data to the growing body of research on syntax and semantics of these expressions holds a great deal of promise.

Moving on to morphology, most Caucasian languages are agglutinative—that much can be deduced from the examples presented so far. Northwest Caucasian and Kartvelian languages are characterized by long verb forms that include multiple indexing of person and number of participants, aspect, *Aktionsart*, and applicative verbal

affixes. Such complexity of verb forms, coupled with extensive pro-drop, has led researchers to characterize Northwest Caucasian and Kartvelian languages as polysynthetic (Testelets, 2009a; Wier, 2011).[17]

Indexical shift is another structural phenomenon common to the area. Indexicals are expressions that depend on the context of utterance (e.g., *I, you, now, here, tomorrow*). Traditional accounts of indexicals assume that their referents are fixed regardless of the syntactic environments they are used in. Therefore, indexicals always refer to the actual context of utterance (Kaplan, 1989; Sudo, 2012). Over the last two decades, researchers have shown that in a number of languages, indexicals may be interpreted in the context of the utterance (direct reading), or in relation to the reported context (the shifted reading). In the Georgian example in (4), the first person pronoun is ambiguous; it can either refer to the speaker or to Nino. Referring to the speaker, the indexical receives its standard, unshifted interpretation based on the actual context of the utterance. Referring to Nino, the same expression is interpreted in the context of the report.

(4) Georgian
   nino-m   tkv-a   (rom)   xval   mo-val-o.
   Nino-ERG   say-AOR.3SG   that   tomorrow PV-go.FUT.1SG-QUOT
   'Nino said that I [=the speaker] will come tomorrow.'
   'Nino said that she will come tomorrow.'

Aside from Georgian, indexical shift has been observed in Svan and Laz (Demirok & Öztürk, 2015; see also chapter 21). It is widely attested in Nakh-Dagestanian (chapter 3; chapter 21; Polinsky, 2015a) and may also exist in Northwest Caucasian languages (Ershova, 2013). Because of this widespread presence, the Caucasus is a promising area for studying indexical shift. However, as with word order or complex consonantal systems, indexical shift is unlikely to be specific to the Caucasus. Kaplan used to describe shifted indexicals as monsters; once the first monsters were uncovered (Schlenker, 1999, 2003), more monsters have been found all over the world (see Deal, 2018, for a recent tally).

So far, the data presented in this section make us think that parallels and similarities across different families in the Caucasus are more or less accidental. The reasons for this may be twofold: first, the languages are indeed diverse and share little beyond basic properties (pro-drop, head-finality); and second, the level of comparison is too coarse-grained, and the features we examine may need to be refined. Below, in no particular order, are some less general properties that appear across the languages of the major families with some recurrence. The list is not exhaustive; rather, it is the beginning of a tally which will hopefully grow as we learn more about the languages of the area.

---

[17] Much in that characterization depends on the criterial properties of a polysynthetic language (see Baker, 1996, for an extensive list): is the indexing of arguments on the verb and extensive pro-drop enough? Is noun incorporation a necessary condition? Answers may be pending but, the characterization of Northwest Caucasian and Kartvelian languages as polysynthetic has thus far led to interesting comparisons of these languages to such polysynthetic exemplars as Salish, Iroquoian, or Algonquian (Testelets, 2009a; Lander & Testelets, 2017).

Furthermore, as with all overviews, certain things have been omitted. For more on the features shared across languages of the Caucasus, see Chirikba (2008b), Klimov (1978), and further references therein.[18]

A morphological optative—the modal form that expresses wishes, desires, potentialities, or hopes—is found in almost all of the area's languages. Example (5) highlights Ancient Greek to illustrate another common property of morphological optatives: co-occurrence with a particular tense-aspect, in this case, aorist:

(5) Ancient Greek
génoitó   moi   katà   tò   rhêmá   sou.
happen.OPT.AOR 1SG.DAT according DET   word   2SG.POSS
'May it happen to me according to your word.'

Optative meaning can be expressed by a number of constructions, but the use of dedicated morphology to do so is quite rare. In the Caucasus, morphological optatives are extremely widespread (Chirikba, 2008b; Dobrushina, 2011; Dobrushina, van der Auwera, & Goussev, 2013).[19] Consider examples from the three indigenous families, as well as some other languages of the area (and see Dobrushina, 2011b, for more examples from the Nakh-Dagestanian family):

(6) a. Adyghe
         qǝ-š'-ere-ç'ǝx   qeʁaǯe-xe-r.
         DIR-LOC-OPT-grow flower-PL-ABS
         'Let flowers grow here!' (Kuznetsova, 2009, p. 291)

     b. Georgian
         man   unda   gadac̣eros   es   c̣eril-i.
         3SG.ERG MOD   3SG.copy.OPT   DEM   letter-NOM
         'He needs to copy this letter.' (Cherchi, 1997, p. 260)

     c. Lezgian
         wa-z   allah-di   hamišan   üsret   gu-raj.
         2SG-DAT Allah-ERG always   help   give-OPT
         'May God always help you.' (Haspelmath, 1993, p. 151)

     d. Kumyk
         tez   jaz   bol-ʁaj   e-di.
         soon summer be-OPT   AUX-PST
         'I wish summer would come soon.' (Dobrushina, 2011b, p. 104)

     e. Judeo-Tat
         soχ-o-m.
         do-OPT-1SG.PST
         'Let me do it!' (chapter 13, p. 610)

---

[18] See also chapters 1 and 3 for a discussion of properties shared across Nakh-Dagestanian languages.
[19] Chirikba (2008b, p. 52) refers to this category as the "potential."

Another common property of languages of the Caucasus has to do with vestiges of a vigesimal counting system found across all three families (Klimov, 1978, pp. 20–21). Comrie (2013) shows that languages of all three indigenous families have a hybrid decimal-vigesimal system in which, "the numbers up to 99 are expressed vigesimally, but the system then shifts to being decimal for the expression of the hundreds, so that one ends up with expressions of the type *x100 + y20 + z*." Given the intensive contact in the area, this is not surprising—the counting systems were shared and could spread from one group to the others.

Unusual argument mapping of objects in a subset of transitive verbs that denote physical contact is another recurrent feature in at least Nakh-Dagestanian and Kartvelian. The verbs in question most commonly include 'hit', 'shoot', 'touch', 'kiss', 'wipe', 'comb', 'paint', and 'stab'. They presuppose an object that is affected by the action, and the medium (instrument) of the respective action. In more familiar languages, the entity undergoing such eventualities is expressed as a direct object, and the medium/ instrument, if expressed at all, is in an oblique form. Yet in Nakh-Dagestanian and Kartvelian languages, the mapping of non-subject arguments appears reversed: the instrument of the action is expressed as a direct object, and the undergoer appears in the dative or locative form (Klimov, 1978, pp. 58– 59).[20] For example:

(7)   Georgian
    gogo-m   k'at'a-s   (top-i)   esrola.
    girl-ERG   cat-DAT   gun-NOM throw.AOR.3SG
    'The girl shot (lit. threw the rifle to/at) the cat.'

(8)   Tsez
    čanaqan-ä   zey-qo   (tupi)   caƛi-n.
    hunter-ERG   bear-POSS.ESS rifle.ABS.IV throw-PST.NWIT
    'The hunter shot (lit. threw the rifle at) the bear.'

Since the expression of the instrument/medium can be omitted, one could form an impression that such verbs are somehow special, missing a direct object entirely—which they are not.

---

[20] Klimov (1978, p. 59) suggests that the same unusual mapping is found in Northwest Caucasian languages, but this observation is not supported by the empirical data. The examples listed in Klimov (1978) represent intransitive verbs whose subject is in the absolutive, whose undergoer is expressed as an indirect object, and whose instrument appears either in the instrumental form or as another indirect object. For example, in (i), the subject is in the absolutive, and the agreement on the verb reflects an intransitive pattern; the instrument is expressed by a PP (*če* is the instrumental postposition that requires an oblique complement), and the notional object is in the oblique form:

(i)   Adyghe
    c_wəweč'ə-m-če   *c_wəwe-r*   *c_wə-me*   *ja-we.*
    rod-OBL-INS   whacker-ABS   bull-PL.OBL   3PL.IO+OBL+DYN-beat.PRS

   'The whacker is racing the bullocks with a whip.' (Arkadiev, Lander, Letuchiy, Sumbatova, & Testelets, 2009, p. 54, glosses modified from the original)

Yet another property shared by languages of the area has to do with the expression of motion events. Talmy (1975, 1985) contends that in the domain of motion events, languages fall into two major types: Path (or v[erb]-framed) languages, which lexicalize the path of motion in the verb and express the manner of motion, if specified at all, outside the verb; and Manner (or s[atellite]-framed) languages which lexicalize the manner of motion in the verb and express the path in a complement ("satellite") to the verb. Romance languages are a common example of the Path type, and Germanic languages instantiate the Manner type. Compare the contrast between Spanish and English in example (9):

(9)   a.   Spanish
           La   botella   entró   a   la   cueva   (flotando).
           the   bottle   entered   at   the   cave   floating
      b.   The bottle floated into the cave.

Although no languages of the Caucasus are exclusively of the Motion or Path type, the Path type is preferred. The manner of motion is rarely expressed by a single verb; instead, we find basic motion verbs such as 'go' or 'come' combined with a nonfinite verb form or an adverb expressing a concomitant action (running/in the running manner, floating/in the floating manner, etc.), as illustrated in (10):

(10)   Chirag Dargwa
       c:ade   š:a   duc'-b-ulq-le   arg-an-de.
       woman+PL.ABS home   run-H.PL-IPFV-CVB   go:IPFV-PTCP-PST
       'The women were running home.'

Furthermore, a number of languages of the area lack such verbs as 'fly' or 'swim'. Taken together, these lexical observations (which have not been systematized so far) are indicative of a promising area of research, one that would combine careful descriptive work on verbs of motion in languages of the area with further testing of Talmy's initial hypothesis.

I have already mentioned the rich morphological makeup of verbs in the languages of the three indigenous families. In particular, most languages allow the construction of morphological causatives of transitives (and further valency increases are also possible, leading to pluritransitive verbs). Throughout the Caucasus, in causatives of transitives, the causer appears in the ergative, the object of the transitive remains in the absolutive, and the causee appears in an oblique form; the alignment where the causee is expressed as the direct object is unattested (Klimov, 1978, p. 57). To illustrate[21]:

---

[21] Northwest Caucasian languages have an extremely impoverished inventory of morphological cases, almost as a mirror image of their case-rich neighbors in the northeast. In Adyghe, the ergative and oblique case have the same exponent, *-m*. Some researchers use that syncretism as evidence that the case is all the same (see chapter 9; Testelets, 2009a). However, the distribution of

(11) Adyghe
    a. *č'ale-m*   *ʁ̩ʷəč'ə-r*   *j-e-wəfe*.
       young.man-ERG iron-ABS 3SG.ERG-DYN-bend.PRS
       'The young man is bending iron.'

    b. *pŝaŝe-m*   *č'ale-m*   *ʁ̩ʷəč'ə-r*   *r-j-e-ʁe-wəfe*.
       girl-ERG   young.man-OBL   iron-ABS   OBL-3SG.ERG-DYN-CAUS-bend.PRS
       'The girl is making the young man bend iron.' (Letuchiy, 2009a, p. 377)

(12) Georgian
    a. *švil-ma*   *p'ur-i*   *mo-i-t'an-a*.
       child-ERG   bread-NOM   PV-VERS-bring-AOR.3SG
       'The child brought bread.'

    b. *deda-m*   *švil-s*   *p'ur-i*   *mo-a-t'an-in-a*.
       mother-ERG child-DAT   bread-NOM   PV-VERS-bring-CAUS-AOR.3SG
       'Mother made the child bring bread.'

(13) Tsez
    a. *kid-b-ä*   *magalu*   *b-aħ-xo*.
       girl-TH-ERG bread.ABS   III-bake-PRS
       'The girl is baking bread.'

    b. *eniy-ä*   *kid-be-q*   *magalu*   *b-aħ-er-xo*.
       mother-ERG girl-TH-POSS.ESS   bread.ABS   III-bake-CAUS-PRS
       'The mother is making the girl bake bread.'

Although this alignment of causatives of transitives is not unique to the Caucasus (it is found in morphological causatives in Japanese, see Harley, 2008), the pervasiveness of this feature among languages of the area is striking. It is found in Ossetic as well (see chapter 14), which suggests that it may be an areal feature.

It is more common to discuss categories and properties present in a given language rather than focus on what is absent. However, some significant "omissions" in the structures of languages of the area should also be noted. In particular, Kartvelian and Northwest Caucasian languages lack infinitives. Instead, they use deverbal nouns (often described as *masdars*, the Arabic term for a verbal noun) or other nominalized forms, such as the supine in the Northwest Caucasian family (Klimov, 1978, pp. 18–19, 78).

With the exception of Armenian (see chapter 13), Old Georgian, and the Northwest Caucasian family (see chapters 9 and 10), Caucasian languages lack articles. That makes them good candidates for testing hypotheses concerning differences in the fundamental design of DP and NP languages (Bošković, 2008), an issue that Öztürk and Eren take up

---

m-marked forms and their control of verbal agreement vary by structural position. Here I adopt the view that -*m* can mark different cases and that the case of the cause in (11b) is oblique, not ergative (see also chapter 18).

in a separate chapter in this volume (chapter 19). Further work in this domain is imperative.

In their demonstrative system, Caucasian languages all distinguish between at least three deictic categories: close to the speaker (*hic*), close to the hearer (*iste*), and away from both speech participants (*ille*). Actual realizations may vary from language to language (Klimov, 1978, pp. 19–20, 83) and often include the distinction between what is visible (*here, there*) and what is out of sight (*yonder*), as well as distinctions based on the position of the reference point on a vertical (*higher*, *lower*, *at the same level/next to*). The three-way distance contrast is also common in locative expressions. Additionally, most of these languages lack dedicated third person pronouns and use demonstratives instead. Given the dizzying array of demonstratives, it would be intriguing to find out which particular items in the demonstrative class are chosen to denote third person referents. Is it 'this', 'that', 'next to the speaker', or 'below the speaker's reference point'? A number of options are attested, and a study that could systematize the use of demonstratives for third person referents across languages of the area is gravely needed.

I.3. Scholarship on Languages of the Caucasus

The data on many languages of the Caucasus are descriptively rich, though not always easily accessible. In order to appreciate the existing scholarship, one must be able to read a series of languages. The earlier research was written up in German, Russian, French, and Georgian, and most of the contemporary literature is in English and Russian.

Early work on languages of the Caucasus can be roughly divided into the work done by local researchers and the work done by outsiders (Klimov, 1986, p. 25). Of the former, most studies were done in Georgia, with an emphasis on Georgian in general and on Bible translations into Georgian in particular. Early local scholars often downplayed the role of other Kartvelian languages. For instance, Megrelian was characterized as a non-standard, uneducated variety of Georgian.[22]

Of the work done by outsiders, early studies on languages of the area are associated with the names of explorers, military officers, and administrators who traveled to the Caucasus and helped map out the area's ethnic and linguistic diversity. The first lexical lists and dictionaries of indigenous languages appeared in the late 1700s (Güldenstädt, 1787–1791; Klaproth, 1812–1814, 1814). More detailed and varied work soon followed. Marie-Félicité Brosset's long and illustrious career studying Georgian and Armenian paved the way for serious historical and philological work in the South Caucasus. Franz Anton Schiefner, Adolf Dirr, and Baron Peter (Pëtr) von Uslar laid the foundations of modern study of Caucasian languages for the three indigenous families. They were not linguists

---

[22] See also chapter 12 for some discussion of this issue.

by training, and their interests spanned ethnography, folklore, history, and language. Thanks to their dedication, we now have detailed grammars and dictionaries of several languages from the area (Dirr, 1904, 1905, 1908, 1928a, 1928b; Uslar, 1887, 1888, 1889, 1890, 1892, 1896, 1979).[23] Baron von Uslar was also responsible for the creation of early Cyrillic-based orthographies for Nakh-Dagestanian languages.[24] The Russian-language journal "Sbornik materialov dlja opisanija mestnostej i plemjon Kavkaza" (SMOMPK) was published in Tbilisi from 1881 through 1915 (additional issues appeared in 1926 and 1929) and remains a valuable resource of ethnographic and linguistic observations. (In fact, many of SMOMPK issues are listed in the references to this *Handbook*.)

Before he gained notoriety for the idea that all of the world's languages descend from a single proto-language with four exclamations as its entire vocabulary, Nicholas (Nikolay/Nikolai) Marr carried out important work on Georgian and Armenian philology. Nikolai Trubetzkoy conducted phonetic/phonological and comparative analysis of languages in the North Caucasus, and his work is still valid and current (e.g., Trubetzkoy, 1922, 1930, 1931). Several outstanding Russian linguists worked in the area in the 1930–1960s, with Moscow, Leningrad, and Tbilisi being established centers of research in Caucasian languages (the first department of Caucasian Language Studies was established at Tbilisi State University in the 1930s). Descriptions of languages produced in these centers remain authoritative sources of data to this day, and sometimes constitute a baseline which allows us to compare an earlier stage of a particular language to the way it is spoken now. Evgeny and Anatoly Bokarev, Arnold Chikobava, Zeynab Kerasheva, Ketevan Lomtatidze, Georgy Rogava, Akaki Shanidze, Nikolay Yakovlev, Lev Zhirkov, Varlam Topuria, Ilia Tsertsvadze, Bakar Gigineishvili—these are just some of the illustrious names on the roster of Caucasologists who worked in Russia/the USSR in the 20th century.

A new model of language study and description was pioneered by Alexander Kibrik and Sandro Kodzasov who, over two decades, led groups of researchers on annual fieldwork trips in the Caucasus. Kibrik's work was undergirded by the desire to combine rigorous theoretical analysis with thorough description of a language (preferably under- or un-described) through intensive fieldwork, typically conducted by entire research teams (see Kibrik, 1972, 1977c, for the main principles of such team fieldwork). Not only did Kibrik and Kodzasov's fundamental work lead to excellent descriptions and analyses of Caucasian languages (A. E. Kibrik, 1977a, 1977b, 1977c, 1992, 1996; Kibrik & Kodzasov, 1988, 1990; Kibrik, Kodzasov, Olovjannikova, & Samedov, 1977a, 1977b; Kibrik & Testelets, 1999; Kibrik, Kazenin, Lyutikova, & Tatevosov, 2001), but it also set a precedent about the importance of group fieldtrips, which serve as incubators for training students and collecting data in all kinds of languages. The Adyghe collection referenced throughout this volume (Testelets, 2009a) is the result of one such field trip.

---

[23] See chapters 3 and 9 for further discussion of early linguistic work in this area.

[24] Russian scholars in the 1920s and 1930s built on that work, creating more alphabets, first based on the Latin script, and later on, as the USSR went back to more imperial aspirations, based on Cyrillic. Nikolay Yakovlev and Lev Zhirkov developed writing systems for a number of Caucasian languages (Alpatov, 2017).

A good place to start for English sources is a special issue of *Lingua* edited by Helma van den Berg (2005a) that includes an overview of each family's phonology, morphology, and syntax. Wixman (1980) provides an excellent ethnographic and sociolinguistic overview of the North Caucasus. Greppin (1989–2004) is a collection of more detailed descriptions, with an overview of each family and descriptions of their languages. Chumakina (2011a) provides a useful annotated bibliography of the main readings on languages of the area, with basic readings for all of the families. Comprehensive bibliographies on particular language families are also available: see Jaimoukha (2009) for Northwest Caucasian; Alekseev and Kikilashvili (2013) and Erschler (2014a) for Nakh-Dagestanian (in Russian and in English, respectively). For Kartvelian, there is no single publication with a relevant bibliography, but the following papers and books have extensive bibliographies: Boeder (2005), Greppin and Harris (1991),[25] and Tuite (1998a).

Fieldwork in the Caucasus is changing. The area is more open to international researchers than ever before, which has led to worldwide collaboration among scholars (Bond, Corbett, Chumakina, & Brown, 2016; Chumakina, Brown, Corbett, & Quillam, 2007a, 2007b), nascent experimental work on languages of the area (Gagliardi, 2012; Lau et al., 2018; Polinsky, Gomez-Gallo, Graff, & Kravtchenko, 2012), and extensive new grammars (Forker, 2013b, is a recent example; see also chapter 3 for more detail). Furthermore, in the late 1990s and early 2000s, the Max Planck Institute for Evolutionary Anthropology in Leipzig supported the publication of dictionaries, language descriptions and documentation, and folklore collections, primarily from the Nakh-Dagestanian family. There is a new sense of urgency in studying the languages of the Caucasus because many have become endangered, due either to dwindling populations or to speakers moving away to areas where Russian or Georgian takes over (see chapter 2; also van den Berg, 1992).

I.4. Structure of This *Handbook*

This *Handbook* is an attempt to bring the descriptive riches of the Caucasus to an English reader, with an additional emphasis on the theoretical promise held by languages of the Caucasus. With that goal in mind, several chapters in this *Handbook* conclude with a section on outstanding issues or topics for future study.

As previously mentioned, the reader who is looking to learn more about the history of languages of the Caucasus may have to look at other references; the emphasis in this volume is on synchronic description.[26] Likewise, someone seeking information about extinct languages that were spoken in the area, for example, Hurrian or Hattic, will be disappointed; this *Handbook* does not include any such descriptions.

---

[25] This volume is part of the series Greppin (1989–2004).
[26] However, chapters 11 and 13 briefly discuss some aspects of the history of Kartvelian and Indo-European languages, respectively.

Part I includes chapters that present a general overview of the area, with emphasis on geography, demographic trends, and social aspects of language use. Demographic research in the Caucasus is still uneven; chapter 2 by Konstantin Kazenin is concerned only with the northern part of the area, and we have been unable to secure comparable chapters for the Kartvelian area—a clear indication where future work is needed.

Each of the indigenous families is described in an overview chapter, and there is also an overview chapters on the local Indo-European languages (Parts II–V). In addition, this *Handbook* includes chapters on selected languages from the main families. Thus, each overview chapter is accompanied by a chapter (or several chapters) on selected languages; special effort was made to include lesser-described languages. For example, in Part IV, the Kartvelian overview is accompanied by a chapter on Megrelian, which has received less attention than the largest language of the family, Georgian (for descriptions of Svan, another understudied language of the family, see Tuite, 1998b, 2018, and references therein).

The Indo-European languages of the Caucasus share striking areal features (see chapter 13). On the contrary, the Turkic languages of the Caucasus do not appear to have attained features specific to the area and present clear examples of Turkic (and broader, Altaic) typology, including vowel harmony and consonantal restrictions at the beginning of a word, the nominative-accusative alignment, and visible agglutination. The relevant languages have been described relatively well, and the interested reader should consult Schönig (1998) for Azerbaijani, Berta (1998) for Kumyk and Karachay-Balkar, and Csató & Karakoç (1998) for Noghay, with further references therein. Since these languages use Cyrillic (see chapter 1), their transliteration conventions are included in Appendix II.

Chapters on language families and individual languages follow more or less the same format, with some deviations. For example, non-finite forms play a crucial role in Nakh-Dagestanian grammars but are much less relevant for the other two families, so the description of such forms is much more extensive in the Nakh-Dagestanian chapters. The discussion of grammatical relations may be more important just for some languages, where their status has been subject of dispute, and may be absent from other chapters where the data are insufficient, or the issue does not even arise. For some languages, certain structural domains are studied comparatively less; while descriptive gaps may constitute obstacles for research, they also offer opportunities for future work.

While the authors of overviews and related language chapters made a concerted effort to coordinate their presentations to avoid duplication, some repetitive material is inevitable, and it may be less repetitive than it seems. For instance, the overview chapter on the Northwest Caucasian family includes charts showing consonant of Abaza and Abkhaz (chapter 9), and so does the chapter by Brian O'Herin (chapter 10). However, the charts represent different dialects, and further still, the authors of the respective chapters have somewhat different views on the sound systems under consideration—an inevitable circumstance in the field, where discoveries are still being made and analyses are being actively worked out. Above I already brought up different views on ergativity in Kartvelian, which are reflected in individual chapters.

I have also mentioned the complex nominal forms in Nakh-Dagestanian languages used to encode spatial meaning. Some researchers analyze them as postpositional phrases (see chapter 3; Comrie & Polinsky, 1998), while others treat them as part of the nominal case paradigm (in chapter 5, Victor Friedman presents arguments in favor of this approach to Lak spatial forms).

The final part of this *Handbook* (Part VI) includes overview chapters that address particular aspects of language structure, from phonetics and phonology to grammar and information structure. The choice of topics was, to a large extent, motivated by available research (and researchers). For instance, there is virtually no research on lexical semantics in languages of the Caucasus and only very preliminary work on propositional semantics of these languages (mainly by Sergey Tatevosov and co-authors, see the chapters on semantics in Tsakhur and Bagvalal descriptions: Kibrik & Testelets, 1999, and Kibrik, Kazenin, Lyutikova, & Tatevosov, 2001, respectively)—that explains one of the gaps in the *Handbook*. It would not be hard to find other areas of inquiry that are missing, and it is my hope that this volume will stimulate new research to fill in these holes.

And finally, some housekeeping notes are in order. Despite its relatively small geographic area, the Caucasus features a nearly overwhelming variety of language names (see also footnote 3). Throughout this *Handbook*, language names have been unified; Appendix I lists the most commonly used names of languages and language groups together with the existing alternatives. For instance, the *Handbook* uses the name *Batsbi* throughout, and Appendix I gives its alternate names: *Bats, Batsaw, Tsova, Tsova-Tush*.

Names in the Caucasus are often more than names; some evoke the history of strife, divisions, or oppression—or other strong connotations. For example, the name *Kartvelian*, commonly adopted for one of the families, is rejected by the Laz, whose language belongs to that family, but who insist on the name *South Caucasian* (see chapters 19 and 22). And the language name *Adyghe*, widely used in the typological literature, and throughout this volume, may be less appropriate than *West Circassian,* the term used in the literature as well (e.g., Smeets, 1984); see chapter 9 for more discussion.[27] While this *Handbook* has adopted a fairly conservative approach, keeping most names as they are found in the bibliographic tradition, it is incumbent upon researchers working in the Caucasus to be cognizant of ethnic or local names going forward.

The variety of spellings and orthographic conventions is yet another issue that any intrepid researcher of the area has to face. With the exception of Azerbaijani, no language in the Caucasus uses Latin script (and many languages do not have writing systems, see chapter 1). Coupled with the complex sound systems, this creates serious challenges in transliterating names of languages or dialects, place names, or names of historical figures and local researchers. Difficulties are further confounded by the existence of several romanization systems for Cyrillic (which is widely used throughout the

---

[27] The choice of names for the Circassian languages is further complicated by aspirations of terminological symmetry; if we use *West Circassian* for Adyghe, it is more appropriate to refer to Kabardian as *East Circassian*. And if we want to keep the more-common *Kabardian*, that may tip the scale in favor of *Adyghe*.

Caucasus) and for Georgian. Appendix II serves to show the most common correspondences between Cyrillic, Latin, and the International Phonetic Alphabet (IPA), which should help with future reading of particular texts.

As much as possible, the authors have tried to use consistent romanization of personal names and names of locations, but old habits die hard and some chapters may have slightly varied transliteration for personal names and names of locations in the Caucasus. This is particularly evident with the romanization of Georgian where several systems compete (the most recent of those is the National System established in 2002 by the State Department of Geodesy and Cartography of Georgia and by the Institute of Linguistics of the Georgian Academy of Sciences). One of the main points of divergence has to do with the representation of ejectives: should they be marked with an apostrophe, with a dot under the consonant symbol, or by capitalization? (This *Handbook* adopts the former convention.) Differences in transliteration of personal names and local names linger, but we have attempted to keep the transcription of the Georgian data as uniform as possible throughout the volume; most exceptions have to do with the transliteration and glossing lifted from earlier work.

The transliteration of Cyrillic follows the scholarly (academic) system (in particular, using the symbols *č, š, ž* among others), and this is used systematically for examples from Russian or the transliteration of book or article titles. Maintaining the same consistency in the transliteration of last names and names of locations is harder, since some names have already been used in a different transliteration; for some, we even find two different spellings (for example, *Testelec* and *Testelets*, or *Daghestan* and *Dagestan*). Where possible we have tried to present the most common transliteration found in the literature; for example, the capital of Georgia is most commonly written in Latin characters as *Tbilisi* (as opposed to the previously used Georgian name *T'pilisi* or the older Russian name *Tiflis*, based on the older Georgian name), and this former name is used throughout this *Handbook*. An additional problem arises when Georgian names appear in a Russian-language source; in such cases, we opted to transliterate the Russian form, for example, Dzheyranishvili (1971, 1984). In the bibliography to the volume the reader may find alternative transliterations of some last names, with a cross-reference to the more common transliteration (for example, *Cagareli*—see *Tsagareli*).

A note on glossing is in order as well. For languages as complex as languages of the Caucasus, morphological division and glossing is an art in and of itself, and a number of conventions have been established for particular languages or families. For instance, infixation is indicated with angle brackets; clitics and affixes are sometimes differentiated by using + and the hyphen, respectively. In Nakh-Dagestanian, where gender agreement is pervasive, Roman numerals are used in glosses to indicate the gender of a noun and the matching of that gender on the agreeing constituent.

A number of glossing abbreviations conform to the Leipzig Glossing rules, but quite a few are not on the Leipzig list—and the list of abbreviations in the beginning of this *Handbook* is understandably long. As with other aspect of data representation, the authors have tried to make the glossing as consistent as possible. Yet some differences are unavoidable, and they go beyond pure terminology. For example, some authors

make a distinction between the generic evidential (EVID) and non-evidential (NEVID): the respective forms express different ways in which evidence was acquired and related to the assertion (was it the event itself that was sensed or was it some other state of affairs that implies the event). Meanwhile, other researchers, in particular those working on descriptions of several Nakh-Dagestanian languages, maintain the more fine-grained distinction between witnessed (WIT), a subtype of direct evidential, and non-witnessed (NWIT), a subtype within the non-evidential category. Accordingly, both categories and the respective abbreviations appear throughout this volume.

ACKNOWLEDGMENTS

GENERAL OVERVIEW OF THE CAUCASUS

LANGUAGES AND SOCIOLINGUISTICS OF THE CAUCASUS

NINA DOBRUSHINA, MICHAEL DANIEL, AND YURI KORYAKOV

1.1 INTRODUCTION

In this chapter, we give an overview of the classification and sociolinguistic situation of the languages of the Caucasus. Section 1.2 presents a summary of family affiliation and classification problems. Section 1.3 discusses language usage statistics as provided by official sources. Section 1.4 provides information on writing systems and the (recent) history of alphabetization. Section 1.5 is an overview of multilingualism in the area. Finally, section 1.6 is a brief discussion of language contact, providing several examples of contact-induced change in the area. As all of us have a deeper knowledge of Nakh-Dagestanian (East Caucasian) languages, this chapter provides more coverage of these languages as compared to languages of the other families.

1.2 FAMILY AFFILIATION

The Caucasus is home to three indigenous language families: Kartvelian (also known as South Caucasian), Northwest Caucasian (other names: Abkhaz-Adyghe, West Caucasian), and Nakh-Dagestanian (also known as East Caucasian or Northeast Caucasian). The three families are often grouped under the rubric "Caucasian languages." There is a considerable representation of Indo-European and Turkic language families, and small ethnic groups speaking Neo-Aramaic (Semitic).

Bold is used for primary language names. *Italics* are used for autonyms, that is, the name used by a given group to refer to their language, as well as for names in languages other than English. Together with the primary language name, the following information is given in parentheses: alternate names, autonyms, and (estimated) number of speakers.[1] The counts do not include speakers outside the Caucasus.

1.2.1 Caucasian Languages

All Caucasian languages were traditionally spoken in southernmost Russia, Georgia, and Azerbaijan (Koryakov, 2002). The only exception is Laz, spoken mainly in Turkey. Many speakers of other languages (e.g., the extinct Ubykh and Circassian languages) moved to Turkey or the Middle East. Section 1.5 discusses the composition of the three groups.

Peter Uslar in his letters of 1864 was the first to voice the idea that the three indigenous language families may be related, though later he himself expressed serious doubts about that (Tuite, 2008, p. 9). This idea was further taken up by a Georgian historian Ivane Javaxishvili (1950) and his successors, A. Chikobava, V. Topuria, G. Rogava, and K. Lomtatidze. It was Chikobava who coined the name "Ibero-Caucasian." The idea was later criticized by many Caucaseologists, especially by G. A. Klimov (1968, 1969). The relationship between Kartvelian, on the one hand, and Nakh-Dagestanian and Northwest Caucasian, on the other, is now rejected by the vast majority of scholars (Comrie, 2005, p. 1; Tuite, 2008, p. 32).

That Northwest Caucasian and Nakh-Dagestanian may be related to each other was proposed by Nikolai Trubetzkoy (1930) and then followed up in *A North Caucasian Etymological Dictionary* by Nikolaev and Starostin (1994). Their reconstruction has been met with skepticism (Nichols, 1997b; Schulze, 1997a); nonetheless "a considerable number of scholars regard (the) North Caucasian hypothesis as at least an interesting possibility worthy of further investigation" (Tuite, 2008, p. 22).

Long-range comparativists further include Northwest Caucasian and Nakh-Dagestanian (North Caucasian) into an even more controversial macrofamily, Sino-Caucasian (Bengtson & Starostin, 2011), relating it, for example, to Basque, Burushaski, and Yeniseian (Starostin, 2010). Kartvelian, on the other hand, is included into the putative Nostratic macrofamily (Bomhard, 2008; Illich-Svitych, 1971).

Theories linking North Caucasian to Hattian (Ardzinba, 1979; Braun, 1994; Chirikba, 1996, pp. 406–432) or Nakh-Dagestanian to Hurrito-Urartian (Diakonoff & Starostin, 1986) were recently considered in detail by Alexei Kassian (2010, 2011a, 2011b). He concludes that there are no grounds to establish any close relationship, although they could all be members of the putative Sino-Caucasian macrofamily.

---

[1] Estimates are based on census figures, supplemented by village population figures for some minor languages where we have more direct data (Andic, Tsezic, and Dargwa).

1.2.2 Kartvelian

Kartvelian includes four languages, spoken to the south of the Greater Caucasus (see Harris, 1991b, and chapter 11 for an overview of the family). **Georgian** (*kʰartʰuli ena*; 3.5 million), by far the largest language of the family, is mostly spoken in Georgia, with smaller communities in Turkey, Azerbaijan, Russia, and Iran. **Svan** (*lušnu nin*; 30,000) is spoken in the northwest of Georgia and Upper Kodor valley in Abkhazia. **Megrelian** (Mingrelian; *margaluri nina*; 300,000) is spoken in the lowlands of western Georgia and southeastern Abkhazia (see chapter 12 for its description). **Laz** (Chan; *lazuri nena*; 22,000) is spoken primarily in northeastern Turkey, and in one part of the village of Sarpi, run through by the state borders of Turkey and Georgia. Speakers of Svan and Megrelian consider themselves ethnic Georgians and use Georgian as their written language. One consequence is that in Georgia, Svan, Megrelian, and Laz are often considered Georgian dialects. Formerly, Megrelian and Laz were mistakenly considered two dialects of a single language, Zan. Church Georgian is a form of Old Georgian used liturgically by Christian speakers of all Kartvelian languages.

Modern Georgian uses the Mkhedruli version and Church Georgian uses the Asomtavruli and Nuskhuri versions of the original Georgian script. Svan and Megrelian are not officially written except for Megrelian in Abkhazia (see section 1.4); occasionally, the Georgian script is used. Laz is written in the Latin alphabet.

1.2.3 Northwest Caucasian

The Northwest Caucasian family comprises Circassian and Abkhaz-Abaza branches. The extinct Ubykh is transitional between the two groups (Chirikba, 1996, pp. 7–8).

Circassian (*adəɣabzɜ*) is considered a single language by its speakers. This view is maintained in the diaspora. However, in Russia, **Adyghe** (West Circassian; *čaxəbzə*; 115,000), spoken in Adygea and Krasnodar Krai, and **Kabardian** (East Circassian, Kabardino-Cherkess; *qʼɜbɜrdɜj-čɜrkʲɜsəbzɜ*; 505,000), spoken in Kabardino-Balkaria and Karachay-Cherkessia, are officially considered distinct languages. Two literary standards were created in the early 20th century, with dialects divided (sometimes arbitrarily) between them. In Russia, both standard languages use Cyrillic. Circassian is also to a limited extent written outside Russia, where the Latin alphabet is often used for this purpose.

Ubykh (*tʷaxá-bza*) was spoken along the coast of the Black Sea (in the area of modern Sochi). Its speakers moved to the Ottoman Empire in 1864, where they switched to Circassian dialects. The last known speaker of Ubykh died in 1992 in Hacı Osman Köyü, a village near the Sea of Marmara.

Abkhaz (*ápʰsš͏ʷa*; 124,000) has two dialects, Abzhui and Bzyb, spoken in Abkhazia. Several other dialects fully moved to Turkey (e.g., Sadz and Ahchypsow). **Abaza** (*abaza bəzŝa*; 36,600) is spoken in Karachay-Cherkessia. Ashqar, officially considered a dialect of Abaza, and Abkhaz seem to be mutually intelligible, while **Tapanta Abaza** is distinct enough to be viewed as a separate language. For further discussion, see chapters 9 and 10.

1.2.4 Nakh-Dagestanian

Nakh-Dagestanian is a language family of six branches spoken in the eastern part of Northern Caucasus with some communities on the southern slopes of the Greater Caucasus. For some languages (Andic, Tsezic, and Dargwa branches), the figures for speakers are given based on village population rather than on census counts. We do this for the following reasons. Highland villages are usually ethnically and linguistically homogenous (see section 1.4.3), and their populations do not shift to major languages. These groups speak their ethnic language as their L1, as suggested by the census. On the other hand, many villagers have moved to the lowlands, and their presence in the towns may now be higher than in their original villages. Such families tend to lose their ethnic language very quickly, sometimes within the first generation of resettlers, but may continue to indicate their ethnic language as their L1 (Rus. *rodnoj jazyk*, lit. "native language") as a way to express their identity. This makes attempts at evaluations based on censuses unreliable (see Friedman, 2010; Kazenin, 2002a). Counts based on village population may provide more accurate estimates of the number of language speakers.

Dagestanian and Nakh languages were originally considered two separate families. Klaproth (1831) recategorized them as two branches of one family (van den Berg, 2005b; Hewitt, 2004)—hence the family name Nakh-Dagestanian. As it turned out, there are almost no shared innovations that are common to all Dagestanian languages as opposed to Nakh languages (cf. Nichols, 2003, p. 241). It is therefore plausible to view the Nakh branch as a sister to other branches of Nakh-Dagestanian (Forker, 2013b; Koryakov, 2002; but see a different conclusion in Nichols, 2003). Chapter 3 presents an overview of the family.

In the *Nakh* branch, **Chechen** (*nu̯oχči:n mu̯ot:*; 1.3 million) and **Ingush** (*ʁalʁa:j mot:*; 293,000) are grouped together under the name Veynakh. Both languages are written in Cyrillic. See chapter 8 for details.

Batsbi (Bats, Tsova Tush; *bacbur mɔt'*; 500) is an unwritten language spoken in northern Georgia. Speakers of Batsbi identify ethnically as Georgians. The language is severely endangered.

Avar, Andic, and Tsezic languages are sometimes grouped together in a single branch (Avar-Andic-Tsezic). In the Soviet censuses after 1937, these languages were not listed separately, and all speakers were registered as Avars. In the 2002 and 2010 censuses they were again listed separately. In practice, however, ethnic identification was not consistent, so that the figures are unreliable. Below in this section, we provide estimates based on village population, but even these figures are sometimes overestimated because many villagers prefer to keep their original address and registration even after they have moved.

Avar (*maʕarul mac', awar mac'*; 693,000) is represented by several dialects (some of which might be distinct enough to be treated as separate languages) in southwestern and central Dagestan and northern Azerbaijan (Zakatala Avar). See chapter 6 for more details.

Andic languages, spoken to the west of Avar in the middle basin of the Andi Koysu River, include **Andi** (*Gʷan:ab mic':i*; 22,500), **Botlikh** (*bujχałi mic':i*; 7,400), **Godoberi**

(ʁibdiλi mic:i; 3,200), **Karata** (kʼːirλi macʼːi; 11,000), **Tukita** (1,300) (usually considered a dialect of Karata, but see Dobrushina & Zakirova, 2019), **Northern Akhvakh** (ašʷaλi micʼːi; 9,500), **Southern Akhvakh** (8,000 total, 350 in highlands, estimate provided by Indira Abdulaeva, personal communication), **Bagvalal** (bagwalal misʼː; 5,500), **Tindi** (idarab mic:i; 9,300), and **Chamalal** (čamalaldub micʼː; 9,600). Tindi and Bagvalal are close to each other, both geographically and linguistically, as are Botlikh and Godoberi. Some languages show visible divergence even on the level of dialects, as Andi in the villages of Andi, Zilo, Rikvani, and especially Muni and Kvankhidatl.

The *Tsezic* (Didoic) languages, spoken to the south of the Andic languages, in the upper-middle basin of the Andi Koysu River, include **Tsez** (Dido; cezjas mec; 12,300), **Khwarshi** (2,200), **Hinuq** (hinuzas mec; 450), **Bezhta** (Kapuchi; bežλˀalas mic; 6,500), and **Hunzib** (honλˀodos mic; 1,000). Language experts consider some of these figures to be underestimates. **Sagada** dialect (soλˀo; 700 speakers) of Tsez is sufficiently divergent to be considered a distinct language. Similarly, the variety of Khwarshi spoken in Inkhokwari is sometimes classified as a separate language.

The *Dargwa* (Dargi) languages are spoken in the southern central part of Dagestan and include a large range of lects traditionally considered dialects of one language, Dargwa. They are all treated as one language in the censuses. On structural grounds, one may distinguish **Northern Dargwa** (133,000), **Muira** (muirala; 34,500), **Tsudaqar** (cʼudqurla; 30,500), **Kaytag** (Kajtak, Kaytak, Xaidaq; χajdaqˀla; 23,600), **Shari** (1,200), **Tanti** or **Southwestern Dargwa** (Tanti-Sirhwa-Amuq; 13,700), **Usisha-Butri** (7,600), **Kubachi-Ashti** (ʕuʁbugan-išťala; 6,200), **Gapshima** (Hapshima; ħabšila; 2,300), **Chirag** (xarʁnilla kub; 2,000), **Sanzhi-Itsari** (sanǯi-icʼarila; 2,000), **Mehweb** (Megweb, Megeb; meħʷela; 800), and **Amuzgi-Shiri** (ʔaˤmuzʁan, x:eran; 200–400).

All speakers of the Dargwa languages and language varieties speak the standard language, which is closest to the Aqusha and Urakhi dialects of Northern Dargwa. The use of standard Dargwa in the south is more limited. The Mehwebs, forming a Dargwa exclave surrounded by Avars and Laks and being taught Avar at school, are not proficient in the standard language. Chirag is probably the most divergent member of the branch, deep in the south of the Dargwa-speaking area. Various Dargwa varieties are endangered due to migration to the lowlands. For more on Dargwa, see chapters 3 and 4.

The **Lak** language (lak:u maz; 140,000) is spoken to the west of Dargwa; Dargwa and Lak may form a deep-level genealogical grouping. For a language with a relatively high number of speakers, Lak does not show strong dialectal variation. See also chapter 5.

The *Lezgic* languages, spoken in southeastern Dagestan and northern Azerbaijan, include **Archi** (aršat:en čat; 1,500), **Tabasaran** (tabasaran čal; 117,000), **Agul** (Aghul; aʁul čal; 27,000), **Lezgian** (lezgi čal; 546,000), **Rutul** (miχaˤbišdi čˀel; 27,300), **Tsakhur** (Tsaxur, Caxur; cˤaˤχna / jiˤqni miz; 20,000, although this figure may be a strong overestimate because of the massive shift of the Tsakhurs of Azerbaijan to Azerbaijani), **Budukh** (budanu mɛz; 200 speakers), **Kryz** (Jek, Alik, Kryts, Dzhek; ɢricä mɛz; 4,400), and **Udi** (udin muz; 4,900, also in Georgia and among recent migrants to Russia). Udi is exceptional in that it is by far the earliest documented language of the family. It is a

descendant or relative of the ancient **Caucasian Albanian** (Aghwan) language or of its sister (Gippert, Schulze, W., Aleksidze, A., & Mahé, 2008; see also section 1.3).

Khinalug (*kätš mic'*; 2,200) is spoken in northern Azerbaijan. It may be distantly related to the *Lezgic* branch, with which it is traditionally grouped. Today this closeness is sometimes explained by a strong Lezgic influence.

Together with Batsbi, Budukh is one of the few Nakh-Dagestanian languages which seems to be immediately endangered. The village of Budukh, the only village speaking the language, is reported to be shifting to Azerbaijani (Adigoezel Hacijev, personal communication, July 4, 2018). While language shift is widespread in the lowlands, affecting all Nakh-Dagestanian languages, Budukh is one of the few known cases of language shift currently in progress right at the original location where a Nakh-Dagestanian language is spoken. More such cases may have happened relatively recently but remained undocumented, such as a probable shift from Tabasaran to Azerbaijani or Lezgian in some villages in the south of Dagestan (Genko, 2005, p. 203).

1.2.5 Indo-European Languages

Russian is the most widely spoken language of the Caucasus (20.5 million). In terms of the number of L1 speakers, it is slightly behind Azerbaijani (8.3 vs. 8.5 million). Apparently, the number of Russian speakers (both L1 and L2) in the southern Caucasus including Georgia, Armenia, and Azerbaijan has declined since the fall of the Soviet Union, although no reliable statistics are available. The use of Russian in Dagestan, on the contrary, increases. All ethnic groups speak Russian as L1 in towns, where the majority of population lost their ethnic languages. In villages, monolingual speakers of ethnic languages are exceptional, and, in most cases, Russian is the only lingua franca between neighboring villages (see section 1.5). Speakers of **Ukrainian** (110,000) are scattered throughout the Caucasus but are especially dense in northern to central Krasnodar Krai.

Several *Iranian* languages are spoken in the Caucasus. One is a *Northwestern Iranian* language, **Northern Kurdish** (Kurmanji; *kurmanǯi*; 60,000) spoken by scattered communities in Armenia, Azerbaijan, and some parts of Northern Caucasus (especially the Republic of Adygea). The Yezidis (*Êzidî*), a separate Kurdish religious sect, claim their variety of Kurdish to be a distinct language, Ezdiki. It appeared as such in the Armenian census, while the census administration in Russia, after linguistic consultations, merged Ezdiki and Kurdish. **Talysh** (*toliši;* 77,400, possibly more) is spoken in the southeast of Azerbaijan and in adjacent Iranian territory.

Southwestern Iranian is represented by **Tat** (26,600) spoken by three different confessional groups which have their own autonyms for the language: Muslim Tats (*tati*; northeast Azerbaijan and the suburbs of Baku), Mountain Jews (*ǯuhuri*; Quba in Azerbaijan, few speakers in Dagestan, others moved to Israel or scattered over other towns in the Caucasus and elsewhere in Russia; written in the Hebrew script) and the nearly extinct Christian Tats (*pʰarseren*; Armeno-Tat; formerly in Madrasa and Kilvar in Azerbaijan, but moved to Armenia and Russia).

Northeastern Iranian is represented by **Ossetic** (0.5 million), spoken in the Russian Republic of North Ossetia and in the self-proclaimed republic of South Ossetia. Ossetic has two strongly divergent dialects, Iron and Digor (see chapters 13 and 14).

The Indo-Aryan branch of Indo-European is marginally represented by **Romani** (*romani čʰib*; 34,000) dialects, the language of the Roma (Gypsies) scattered throughout the Northern Caucasus and Georgia. In many areas, the Roma have assimilated linguistically to the surrounding languages. Romani is not an official language anywhere in the Caucasus. In recent years, it has had limited use in writing (Cyrillic). Closely related **Domari** (Karachi, Mitrib, Kaloro, Cingāna) is spoken by scattered Dom communities in Azerbaijan (and elsewhere in the Middle East).

Lomavren (Bosha, 35,000–40,000) is a nearly extinct mixed language, spoken by the Lom people in Armenia, southern Georgia, and northeastern Turkey. It has retained most of its Indo-Aryan lexicon, but its grammar is almost entirely Armenian.

The Armenian branch is represented in the Caucasus by a dialect network usually considered to be a single language, **Eastern Armenian** (*hajeren*; 3.6 million). Some dialects lack mutual intelligibility. It is also the main language of Armenia and Nagorno-Karabakh (where a strongly divergent Artsakh dialect is spoken as L1). Eastern Armenian is also spoken in Georgia and Iran. Its sister language, **Western Armenian**, is used in Turkey and by the majority of the Armenian diaspora outside Iran. In the Northern Caucasus, there are two divergent Western Armenian varieties. One is **Nor-Nakhichevan** Armenian, spoken in the Rostov region by descendants of Crimean Armenians. The other is **Hamshen Armenian** (*homšecʰma*), spoken by Christians in Abkhazia and Krasnodar Krai, and also by Muslims in northeastern Turkey, who do not consider their language to be Armenian (Koryakov, 2018). Many Armenian dialects became extinct in the aftermath of the genocide in the early 20th century.

At least two *Hellenic* languages are present in the Caucasus. Divergent varieties of **Pontic** (*roméjka*, 30,000–40,000) are spoken in southern Georgia (most of the speakers emigrated to Greece or Russia), Northern Caucasus, Abkhazia, and Armenia. **Kappadokian Greek** was spoken in Cappadocia (Central Turkey). In the 1920s, most speakers were forced to migrate to Greece, but some moved to Georgia, where Kappadokian Greek is still marginally present in some communities.

1.2.6 Turkic Languages

Turkic languages are widely spoken in some parts of the Northern Caucasus, especially in Azerbaijan, but also in Dagestan.

Turkish (*türkče*; 85,000) is represented in the Caucasus mainly by **Meskhetian Turkish** (*ahɨska türkče*), originally in southern Georgia, and now widely dispersed throughout Russia, Central Asia, Azerbaijan, Ukraine, Turkey, and the USA. **Azerbaijani** (Azeri; *azərbajžan dili*; 9 million) is the main language of Azerbaijan (with more speakers in Iran than Azerbaijan) and a few villages in southern Dagestan, where it is one of the official

languages. **Kumyk** (Kumuk; *qumuq til*; 403,000) is a language of lowland Dagestan; it used to be a lingua franca for some parts of Dagestan. **Karachay-Balkar** (*qaračay-malqar til, tawlu til*; 299,000), the language of two ethnic groups, Karachays and Balkars, occupies the highest areas in Karachay-Cherkessia and Kabardino-Balkaria, respectively. **Noghay** (Nogai; *noɣaj tili*; 73,000) is spoken in several areas in the northern Caucasus, mostly, in northeastern Dagestan and Karachay-Cherkessia.

1.2.7 Semitic

Speakers of several *Northeastern Neo-Aramaic* languages, the largest branch of Neo-Aramaic, migrated to the southern Caucasus in the 19th and 20th century and are scattered throughout Armenia (2,400 speakers), Georgia (several dozen speakers left after mass migration to Russia) and the Russian northwestern Caucasus (1,400 speakers). Today, **Urmian Jewish Aramaic** (*lišān didān, lišānān*) is still spoken in Tbilisi and in the village of Urmia in the Krasnodar Krai (around 80 speakers). Several hundred speakers of **Northern Bohtan Aramaic** (Hértevin; *sôreth*) live in Krymsk and Novopavlosk (southern Russia), where they moved from Gardabani (Georgia).

1.3. Official Statistics on Language Speakers and Users

Only a few of the ex-Soviet republics kept Russian as their official language. In the Caucasus, Armenia, Azerbaijan, and Georgia each have one official language (Armenian, Azerbaijani, Georgian, respectively), as does the unrecognized state of Artsakh (Nagorno-Karabakh Republic), where the official language is also Armenian. The partially recognized republics of South Ossetia and Abkhazia declared Russian as their official language, in addition to the dominant languages (Ossetic and Abkhaz, respectively). The situation with actual language use is, of course, much more complicated.

Almost all of the countries in the Caucasus are successors to the Russian Empire and the Soviet Union, which had one of the world's longest uninterrupted histories of ethnic and linguistic statistics based on systematic census data. The first census in which language data were collected was the Russian Imperial Census of 1897. It was followed by the Soviet "All-Union" census in 1926. The 1926 census included a question regarding the respondent's first language (Rus. *rodnoj jazyk* 'mother tongue'). The question was repeated in Soviet censuses almost every decade: in 1939,[2] 1959, 1979, 1980, and 1989.

---

[2] The census was originally carried out in 1937. It was then announced that its results had been intentionally distorted by the organizers (accused of being "enemies of the people"), and the census was redone two years later.

Table 1.1 Last Two Decennial Censuses in the Caucasus

Country	Year	Year	Census Question[a]
Abkhazia[b]	2003	2011	L1
Armenia	2001	2011	L1, OL
Azerbaijan	1999	2009	L1, proficiency in Azerbaijani, OL
Georgia	2002	2014	L1, proficiency in Georgian
Artsakh[b]	2005	2015	L1
Russia	2002	2010	proficiency in Russian, OL, L1
South Ossetia[b]	–	2015	L1
Turkey	2000	2011	–[c]

[a] L1 – first language, OL – other language[s] spoken fluently.
[b] Unrecognized or partially recognized states.
[c] The 1965 census in Turkey included a question regarding the language usually spoken at home, as well as a question on other languages used by the respondent.

The practice of collecting detailed ethnic and linguistic statistics in censuses continued after the collapse of the USSR (see Table 1.1.). In most cases, this practice was further developed and sociolinguistically refined.

Table 1.2 presents the data on the largest languages of the Caucasus. It includes data from Abkhazia, Armenia, Artsakh, Azerbaijan, Georgia and South Ossetia, and nine administrative units of the Russian areas of the Caucasus.

Table 1.2 Ten Most Spoken Languages of Caucasus (2009–2015 Censuses)

Language	L1	L2	L1+L2
Azerbaijani	8,506,270	524,841	9,031,111
Russian	8,320,492	12,197,152	20,517,644
Armenian	3,510,783	77,281	3,588,064
Georgian	3,310,978	181,416	3,492,394
Chechen	1,284,271	9,473	1,293,744
Avar	618,673	112,853	738,526
Lezgian	532,614	13,786	546,400
Kabardian	488,700	15,672	504,372
Ossetic	462,724	19,756	482,480
English	–	771,422	771,422
Total	**31,182,548**		

In the rest of this section, we present and discuss observations that may be drawn from the census data on language usage, country by country in alphabetical order. Political affiliations of some territories in the Caucasus are vigorously disputed. The authors do not take any political stance on the territorial conflicts of the region.

1.3.1 Armenia

Table 1.3 reproduces the data from the 2011 Census in Armenia.[3] Armenian is the only official language in the country. Russian continues to be widely used, even if its use has considerably reduced in the decades following the collapse of the USSR (see section 1.5 for qualitative discussion). Note the speakers of Neo-Aramaic, whose presence was officially reported as early as in the 1897 Census. Yezidi and Kurdish, though counted separately, are in fact the same language, and the difference in the census data should be attributed to self-identification based on religious affiliation.

Table 1.3 Linguistic Composition of Armenia, 2011

Language	L1	L2
Armenian	2,956,615	43,420
Yezidi	30,973	5,370
Russian	23,484	1,591,246
Assyrian (= Neo-Aramaic)	2,402	1,468
Kurdish	2,030	1,309
Ukrainian	733	1,151
Greek	332	2,136
Georgian	455	6,151
Persian	397	4,396
English	491	107,922
French	–	10,106
German	–	6,342
Other	913	10,339
Refuse to answer	29	–
Total	**3,018,854**	

---

[3] Armenian Census 2011, http://armstat.am/ru/?nid=517

1.3.2 Azerbaijan and Artsakh (Nagorno-Karabakh)

Table 1.4 shows language data from the official 2009 Census in Azerbaijan, including data on Artsakh (Nagorno-Karabakh). For minority languages (i.e., languages other than Azerbaijani, Russian, and English), the counts only include respondents who indicated the corresponding ethnicity. For instance, only those people who declared themselves Talysh were asked about their proficiency in Talysh and, depending on their answer, identified as L1 or L2 speakers of Talysh. All respondents were asked about their proficiency in Azerbaijani, Russian, and English. A comparison of the 2009 and 1999 census data also suggests that the number of L1 speakers is underestimated, at least for Talysh, Georgian, and Tsakhur. For Georgian, apparently only Christian Georgians were counted, while Muslim Georgians were counted as Azerbaijanis, and

Table 1.4 Linguistic Composition of Azerbaijan, 2009[a]

Language	L1	L2
Azerbaijanis	8,148,282	519,417
Lezgians	162,450	2,885
Armenians	120,180	0
Russians	117,988	626,877
Talyshis	47,600	560
Avars	46,463	398
Turks	31,806	684
Tatars	24,139	104
Ukrainians	20,984	22
Tats	19,001	277
Tsakhurs	11,722	111
Georgians	9,682	20
Jews	8,509	55
Kurds	2,202	964
Kryz	1,254	18
Udis	3,773	0
Khinalugs	2,143	2
Other	7,648	228
English	–	71,380
Total	**8,922,447**	

[a] Population by ethnic group: language and L2 proficiency, 2009, http://www.stat.gov.az/source/demoqraphy/en/1_11-12en.xls

Table 1.5   Linguistic Composition of Artsakh, 2005[a]

Language	Native Speakers
Armenian	136,366
Russian	1,274
Ukrainian	7
Other	90
Total	137,737

[a] Preliminary results of the 2015 Nagorno-Karabakh population census: http://www.stat-nkr.am/hy/2010-11-24-10-40-02/597—2015

their proficiency in Georgian was not reflected in the census. Finally, Budukh is not listed in the statistics, while Kryz and Khinalug, not listed in the previous censuses, are included.

Another issue has to do with the population of Artsakh (Nagorno-Karabakh). The figures were calculated based on the data of the 1989 census and thus include the Azerbaijani refugees from the Artsakh territory. Artsakh conducted its own post-Soviet censuses in 2005 and 2015. Table 1.5 provides available linguistic data (on L1).

1.3.3   Georgia, Abkhazia, and South Ossetia

Table 1.6 shows the data from the 2014 census in Georgia.[4] It includes the number of native speakers per language together with the number of those who speak fluent Georgian. Azerbaijanis and Armenians have the lowest number of Georgian speakers. As in the Soviet censuses, recent Georgian censuses do not include statistics for some minority languages: Svan, Megrelian, Laz, and Batsbi (Nakh branch of Nakh-Daghestanian). For these speakers, only their fluency in non-native language(s) is registered. As is evident from a comparison with the 2002 census, Greek, Kurdish, and Kist (a dialect of Chechen), whose speakers are not considered to be ethnic Georgians, are included as "Other" in the 2014 census.

The low counts for Ossetic and Abkhaz indicate that the census does not include data on the breakaway republics of Abkhazia and South Ossetia, which are given in Table 1.7 and Table 1.8, based on the official counts provided by these republics. In Table 1.7, the majority of those who indicated Georgian as their L1 are in fact first-language speakers of Megrelian. South Ossetia conducted its only census in 2015. It included the question on ethnic affiliation, but not on language use (see Table 1.8).

---

[4] 2014 General Population Census Results, http://census.ge/en/results/census

Table 1.6 Linguistic Composition of Georgia, 2014

Language	Native Speakers	Fluently Speak Georgian	Do not Fluently Speak Georgian	Not Stated
Georgian	3,254,852	3,254,852	—	90
Azerbaijani	231,436	43,579	172,134	7,642
Armenian	144,812	57,316	74,258	3,640
Russian	45,920	29,179	9,099	13,238
Ossetic	5,698	4,831	189	15,723
Abkhaz	272	163	19	678
Other	30,742	19,095	8,007	72
Not Stated	72	—	—	—
Total	3,713,804	3,409,015	263,706	41,083

Table 1.7 Linguistic Composition of Abkhazia, 2011[a]

Language	Total Native Speakers	Of the Same Ethnic Group	Of Other Ethnic Groups
Abkhaz	121,697	120,817	880
Armenian	40,831	40,731	100
Georgian	38,020	37,933	87
Megrelian	3,112	3,112	—
Russian	28,580	21,921	6659
Greek	—	862	—
Ukrainian	717	706	11
Ossetic	—	426	—
Abaza	—	308	—
Turkish	—	370	—
Romani	—	253	—
Estonian	—	225	—
Other	7,748	—	3,955
Total	240,705		

[a] Census results for the territory of Abkhazia 1886–2011, http://www.ethno-kavkaz.narod.ru/rnabkhazia.html

As in Table 1.4, the number of minority language speakers in Table 1.7 only represents the respondents who identified with a given language based on their ethnicity. As a result, local patterns of multilingualism (the Abaza speaking Ossetic, etc.) are not accounted for. The only data on multilingualism that are available cover the knowledge of the major regional languages: Abkhaz, Armenian, Georgian, Russian, and Ukrainian.

Table 1.8 Ethnic Composition of South Ossetia, 2015[a]

Ethnic Group	Population
Ossetians	48,146
Georgians	3,966
Russians	610
Other	810
Total	53,532

[a] South Ossetia census results: http://ugosstat.ru

1.3.4 Russia

The exact number of language questions and their formulations vary from census to census. The 2002 census included the following relevant questions: Do you speak Russian? What other languages do you speak? (up to three languages). In the 2010 census, an additional question (taken from the Soviet censuses) was added: What is your first language?

Russian is the only official language of the entire Russian Federation. In some of the republics that form the Russian Federation, local languages are normally co-official languages, written in the Cyrillic alphabet, except indicated otherwise. The area of the Caucasus that is part of the Russian Federation comprises seven federal republics[5]: Adygea (Adyghe), Chechnya (Chechen), Ingushetia (Ingush), Kabardino-Balkaria (Kabardian, Karachay-Balkar), Karachay-Cherkessia (Karachay-Balkar, Kabardian, Noghay), North Ossetia-Alania (Ossetic), and Dagestan. According to the language legislation of the Republic of Dagestan, all indigenous languages are official. These languages are not, however, officially listed in the language policy laws. In practice only the written languages are treated as official: Agul, Avar, Chechen, Dargwa, Kumyk, Lak, Lezgian, Noghay, Rutul, Tabasaran, Tat, and Tsakhur.

Table 1.9 provides combined data on L1 and L2 for nine administrative units of Russia conventionally included in the Northern Caucasus (Krasnodar and Stavropol Krai, Republics of Adygea, Chechnya, Dagestan, Ingushetia, Kabardino-Balkaria, Karachay-Cherkessia, and North Ossetia–Alania). The data are based on the 2010 census. Listed are all languages traditionally spoken in the Caucasus, as well as immigrants' languages with more than 1,000 speakers. The total number given in the last line is the total population of the nine Caucasian administrative units of the Russian Federation, not the result of adding together the numbers of speakers for each language in the table. The latter would

[5]   We show languages spoken in each republic in parentheses following the name of the republic.

Table 1.9 Linguistic Composition of Northern Caucasus (Russia), 2010

Language	L1+L2
Russian	14,350,496
Chechen	1,293,744
Avar	685,726
Kabardian	504,372
Dargwa[a]	456,151
English[b]	429,120
Ossetic	428,636
Kumyk	402,373
Lezgian	363,100
Karachay-Balkar	299,179
Ingush	292,609
Armenian	245,020
Azerbaijani	171,703
Lak	140,394
Tabasaran	116,778
Adyghe	114,970
German[b]	107,346
Ukrainian	75,419
Noghay	73,008
Georgian	57,907
Turkish	37,161
Abaza	36,555
Romani	34,009
French[b]	28,883
Rutul	27,225
Agul	26,953
Tatar	24,534
Greek	23,280
Turkmen	12,063
Tsez	11,994
Kurdish	11,505
Uzbek	10,290
Arabic[b]	10,251

(*Continued*)

Table 1.9 Linguistic Composition of Northern Caucasus (Russia), 2010   (Continued)

Language	L1+L2
Tsakhur	9,364
Belarusian	8,591
Spanish[b]	8,308
Bezhta	5,899
Kazakh	5,712
Andi[c]	5,417
Moldovan	4,149
Korean	3,997
Tajik	3,492
Polish[b]	3,157
Italian[b]	2,801
"Dagestanian"[a]	2,268
Tindi[c]	2,109
Latin[b]	1,941
Neo-Aramaic	1,910
Circassian	1,896
Abkhaz	1,758
Bulgarian	1,756
Khvarshi	1,729
Chuvash	1,686
Mordvin	1,649
"Jewish"[d]	1,447
Bagvalal[c]	1,435
Kyrgyz	1,425
Tat	1,423
Lithuanian	1,215
Bashkir	1,162
Persian	1,059
Archi	961
Hebrew	945
Hunzib	918
Chamalal[c]	470
Udi	433
Megrelian	230

Karata[c]	222
Akhvakh[c]	185
Botlikh[c]	179
Godoberi[c]	120
Total	**15,095,469**

[a] Collective term for more than one language.
[b] A second language for most respondents.
[c] For these languages, the numbers are strongly underestimated.
[d] "Jewish" includes Yiddish and Hebrew (both of which were also listed separately) and Judeo-Tat (Mountain Jewish), which are all genetically unrelated.

be substantially higher, because the figures of speakers per language include not only L1 but also L2 speakers (sometimes respondents indicated several L2s).

Speakers of some languages may identify themselves with another, larger ethnic group and its language. This can lead to an underestimation of the number of speakers for some languages (Table 1.9, note c). For instance, the counts for most Andic languages are much lower than independent estimates. This reflects the fact that many Andic speakers are registered as Avars. Section 1.2 provides more realistic figures (based on village population, usually monolingual in terms of L1). Some of our language consultants report pressure from the authorities to register as one of the majority ethnic groups.

1.4 Writing and Scripts

Not all languages of the Caucasus are written. By 2017, 27 languages spoken in the Caucasus had the official status of a written language. Georgian and Armenian literary traditions have existed continuously from the early Middle Ages. In various other areas, Arabic was the main language of written communication.

Udi (Lezgic branch) (or its sister dialect) was also written in the Middle Ages in a script probably related to Old Armenian and Old Georgian: the so-called Aghwan or Caucasian Albanian. The script subsequently fell into oblivion, and its language, long known only from short texts such as inscriptions on artefacts, has remained a mystery, identified by different scholars with many languages in the Caucasus and beyond. Very recently, after a palimpsest dating to the end of the first millennium was found in a monastery on Mount Sinai and has become accessible for academic research, the language has been reliably identified as an older stage of Udi (Gippert & Schulze, 2007; Gippert, Schulze, Aleksidze, & Mahé, 2008).

Apart from the three languages which have ancient scripts, and apart from the occasional use of the Arabic script for Avar (and several other larger languages), the languages of the Caucasus acquired their writing systems in the late 19th to the 20th century. Modern literatures exist primarily in Azerbaijani, Abkhaz, Chechen, Ingush, Avar, Dargwa, Lezgian, and Lak. The scripts in use changed several times depending on the political situation. For example, the first Abkhaz alphabet was created in 1862 by Peter Uslar and was based on the Cyrillic script. In 1926, it was replaced by an alphabet based on the Latin script. From 1938 to 1954, the Georgian alphabet was used in schools. In 1954, Cyrillic was re-introduced and has been used ever since.[6]

Not all languages that are officially written are equally common in writing. For example, a Cyrillic-based writing system for Rutul was introduced in 1992. Before that, there were almost no written documents in Rutul. Azerbaijani was used for official documents, religious texts, and poetry. After 1992, Rutul classes were incorporated into the school curriculum, two hours a week. Several textbooks have been published. Currently, the local newspaper *Rutul'skie Vesti* ('Rutul News') publishes most articles in Russian and only some in Rutul. However, our consultants report difficulties with reading the articles, because the newspaper uses the variety of Rutul spoken in the village of Rutul (the administrative center), whereas most villages have their own dialects, sometimes considerably different from the Rutul of Rutul Village. Rutul speakers have little motivation to learn another dialect of Rutul to read it. First, the written variety of a different dialect may not align with ethnic identity; and second, all important documents are in Russian anyway. Rutul classes in school are not popular among either students or their parents. The situation with written Tsakhur, a sister Lezgic language spoken in the neighboring villages, seems to be quite similar.

Archi, a minority language spoken in a single village, provides another example of issues related to transliteration. In the 2000s, a team led by Aleksandr Kibrik suggested a practical orthography for Archi. It was used in a number of texts published by Chumakina, Arkhipov, Kibrik, & Daniel (2008), and in the online Archi dictionary compiled by the Surrey Morphology Group (Chumakina, Brown, Corbett, & Quilliam, 2007a). In practice, however, this writing system is only used in interactions between field researchers and Archi native consultants (e.g., when transcribing Archi recordings).

Khinalug, a language spoken in a single village in the north of Azerbaijan, was not written until the 1980s. In the 1980s, a Cyrillic-based orthography was used by the local poet, Rahim Alhas. A slightly different Cyrillic orthography was independently used by Faida Ganieva in her dictionary of Khinalug (Ganieva, 2002). At that time, however, Azerbaijani had already switched to a Latin-based orthography, so attempts to

---

[6] Even for the languages considered unwritten, there is a possibility of discovering relatively old manuscripts in personal archives. For example, Archi, a Lezgic language spoken in one highland village (about 1,500 speakers), has seen the publication of several small poetic religious texts, apparently dating back to the mid-19th century (Magomedxanov, 2009). At that time and earlier, an Arabic script was used.

create Cyrillic orthographies were ill-timed. A systematic Latin-based orthography was first suggested by a team lead by Aleksandr Kibrik and then further negotiated with the community in the 2010s by Monika Rind-Pawlowski. Unlike the situation with Archi, some villagers seem to be enthusiastic about the Khinalug orthography; still, its application and utility seem unclear.

The situation is not much different with major Nakh-Dagestanian languages such as Avar, Dargwa, or Lezgian. It is clear that most Dagestanians read and write Russian much more often (if not exclusively) than their native language, even if they are speakers of Avar or Lezgian. There are similar reports for Ingush. Local enthusiasts use the written minority languages, especially in writing poetry, but the poetry is hardly ever read. The audience of the many newspapers which publish materials in languages other than Russian is very limited. Having said this, no true and reliable estimate of the written use of these languages has been made to date, so that these claims must be considered impressionistic.

The first known attempts to create a script for Megrelian were made in the second half of the 19th century. Since then, however, Megrelian has been written mostly for academic purposes of language documentation, by the early Soviet administration in the 1920s and 1930s, and currently by several newspapers (in particular, in the breakaway region of Abkhazia). In addition, several local enthusiasts in Georgia use this system to write prose and poetry. Native speakers of Megrelian not involved in these activities are unlikely to be fluent in reading the written language (Alexander Rostovtsev-Popiel, personal communication, May 5, 2017).

Some major languages also have practical problems with established writing systems. The first Ingush alphabet (created in 1921) was based on Latin script. In 1938, it was converted to Cyrillic, and that version is still in use. Johanna Nichols (personal communication, May 10, 2017) indicates problems which arise due to the fact that Ingush and Chechen both have schwa-zero alternations, which make it difficult to use a phonemic orthography. Speakers complain that pronunciation and spelling are very distinct. The younger generation, who are generally less literate in Chechen or Ingush, do not write weak vowels, which makes the spelling somewhat chaotic.

Table 1.10 shows the written languages of the Caucasus. The writing systems are divided into traditional (the ones that have been in existence since the Middle Ages), systems that were introduced in the late 19th to early 20 century, transitional systems, and recently established systems.

All modern scripts in the parts of the Caucasus within the Russian Federation are based on Cyrillic. Sound types alien to Russian are represented by combinations of letters rather than by the use of diacritics; especially frequent and multifunctional is the symbol *I* (Rus. *paločka*), primarily used following obstruents to represent ejectives. There are, however, slight differences in use of symbols even among different languages within Dagestan. This stems from the fact that sound inventories differ among languages, so that sound types present in some languages may use character combinations that are used for other sound types in some other languages. This does not create confusion because speakers are usually literate in only one language of Dagestan, and

Table 1.10 Scripts Used by Caucasian Languages

Language	Fam.	Arabic Script	Roman Script	Cyrillic Script	Other Scripts
TRADITIONAL					
Georgian	K				since 5th c. (Georgian)
Armenian	IE				since 5th c. (Armenian)
Aghwan	ND				5th–10th cc. (Aghwan)
LATE 1800s–EARLY 1990s					
Adyghe	NWC	1918–1927 (spor. 19th c.)	1927–1937 (att. 1980s)	1937 (spor. 19th c.)	
Kabardian	NWC	1920–1923	1923–1936 (att. 1980s)	1936–	
Abaza	NWC		1926–1938	1938–	
Abkhaz	NWC		1926–1938	1862–1926, 1954–	1938–1954 (Georgian)
Chechen	ND	1918–1925 (spor. 19th c.)	1925–1938 (att. 1990s)	1938–	
Ingush	ND	1918–1923 (spor. 19th c.)	1923–1938	1938–	
Avar	ND	1918–1928 (spor. 15th c.)	1928–1938 (att. 1990s)	1938–	spor. 10–14th cc. (Georgian)
Dargwa	ND	1918–1928 (spor. 16th c.)	1928–1938	1938–	
Kaytag	ND	(spor. 14th c.)			
Lak	ND	1918–1928 (spor. 15th c.)	1928–1938	1938–	
Lezgian	ND	1918–1928 (spor. 19th c.)	1928–1938	1938–	
Tabasaran	ND		1928–1938	1938–	
Tat	IE		1928–1938, 1990s	1938–	1870s–1928 (Hebrew)
Talysh	IE	Currently (Iran)	1928–1938, 1990s	1938–1990s (spor.)	
Ossetic	IE		1923–1938	(att. 18th–19th cc.) 1844–1923, 1938–	att. 18th–19th cc., 1938–54 (Georgian)

Azerbaijani	Tu	? –1929 (Az.) / Currently (Iran)	1925–1939, 1992–	1939–2001/ Currently (RF)
Kumyk	Tu	19th c.–1928	1928–1938	1938–
Karachay-Balkar	Tu	1910–1925	1924–1938 (att. 1990s)	1938–
Noghay	Tu	19th c.–1928	1928–1938	1938–
TRANSITIONAL				
Megrelian	K			(spor. 1860s) 1920–1933, spor. 1990s, 2003–
Udi	ND		(att. late 1990s)	1935–1936, att. 1990s
Rutul	ND		1928–1938	1938–1940,1992–
Tsakhur	ND		1928–1938 (att. 1990s)	1938–1940,1992–
RECENTLY ESTABLISHED				
Agul	ND			1992–
Andi	ND			att. 1992
Tsez	ND			att. 1993
Laz	K		1984–	

Abbreviations: *att. — attempts*   *spor. — sporadically*

Russian is mostly used in reading and writing anyway. Abkhaz Cyrillic uses a modified version of Uslar's alphabet and is very different from any other Cyrillic alphabet used in the Caucasus.

For transliteration of some scripts based on Cyrillic, see Appendix II.

1.5 MULTILINGUALISM IN THE CAUCASUS

Due to the high language density in the area, multilingualism "was the norm in many Caucasian communities" (Chirikba, 2008b, p. 30). Large ethnic groups in the Caucasus tended to be monolingual, but the language minorities, and in some cases the speakers of major languages living close to linguistic borders, were bilingual in the languages of their neighbors. Under certain socio-economic circumstances, people could also speak

several unrelated languages; this was particularly common among male speakers who traveled and worked outside the home.

The typical linguistic repertoire of Caucasian peoples changed significantly in the course of the 20th century, as command of Russian spread throughout the Caucasus. In the second half of the century, Russian became the language of interethnic communication, and, in many (though not all) places, displaced the use of local L2s. Sometimes Russian even became a danger for the main language of a particular community. However, language death is, so far, not a common phenomenon in the Caucasus.

After the fall of the Soviet Union, the role of the Russian language and culture declined in those parts of the Caucasus that became independent countries: Azerbaijan, Georgia, and Armenia (Pavlenko, 2008). At the same time, the presence of Russian has grown in Dagestan, Chechnya, Adygea, Abkhazia, and South and North Ossetia.

1.5.1 South Caucasus

Before 1991, South Caucasus was part of the Soviet Union and was administratively organized into three first-order units ("Soviet Republics") with five second-order units (Rus. *автономная республика* 'autonomous republic' and *автономная область* 'autonomous district') within them. These were: Georgia (including Abkhazia, South Ossetia, and Adjara), Armenia, and Azerbaijan (including Nagorny Karabakh and Nakhchivan). In 1991, all these territories except Adjara and Nakhchivan declared their independence, but only Georgia, Armenia, and Azerbaijan have been commonly recognized. See Map 1. in this volume.

1.5.1.1 *Azerbaijan*

Before the advent of Russian, Azerbaijanis were essentially monolingual, despite the fact that even in the border areas, bilingualism was (and still is) typical for their neighbors. An Azerbaijani woman who married into a non-Azeri family and lived in a non-Azeri village would hardly learn to speak her husband's language, and most children born into such families would be more likely to speak Azerbaijani as their primary language (Murad Sulejmanov, personal communication, May 8, 2017).

Dialects of Azerbaijani are very much alive. In many places, children can switch from the dialect with their parents and other family members to the standard when communicating with outgroup interlocutors.

In recent decades, Azerbaijani has replaced Russian in many spheres where Russian was dominant during the Soviet era. Russian remains the most widespread second language among middle-aged and elderly Azerbaijanis, but the number of Russian speakers is much less than in the neighboring Dagestan, and young Azerbaijanis usually do not speak it.

Due to the role it plays in politics and media, Turkish is becoming a more popular second language among speakers of Azerbaijani, especially among the youth, since many people watch Turkish TV. We have personally observed the same trend in Azerbaijani villages in Dagestan (e.g., Darvag, Yersi, and others).

Azerbaijani is the dominant language of instruction in primary and secondary schools across the country. Russian is the language of instruction in more than 300 state schools. Most state universities offer undergraduate and graduate programs in Russian, though programs in English are becoming increasingly widespread. In the schools with Azerbaijani as the language of instruction, Russian is sometimes taught as an L2, but English, French, German, Persian, and Arabic are also offered. There is no presence of dialects at schools, and teachers are supposed to teach in the standard language only. To a certain extent, dialects are present in the media, constituting a constant source of worry for the National Council for Television and Radio Broadcast, which will occasionally warn television presenters that they are supposed to restrict themselves to the standard language (Murad Sulejmanov, personal communication, May 8, 2017).

Azerbaijan has numerous linguistic minorities: Armenians, Avars, Budukhs, Georgians, Khinalugs, Kryz, Kurds, Lezgians, Rutuls, Talysh, Tats, and Tsakhurs.

Georgian communities in the Qax region (sometimes called Ingiloys, though Georgians reject this designation as derogatory) remain ethnically homogenous: they resist mixed marriages and have Georgian schools for children, and some locals are even monolingual in Georgian.

Other minority languages are taught as heritage languages in some villages where there are native speakers, but only on an extracurricular basis. The main source of sociolinguistic information on minorities in Azerbaijan is a collection of papers by Clifton et al. (Clifton, 2002, 2003) and a fieldtrip to the Qax region undertaken by two of us in July 2018. We will now discuss the situations of the Talysh, Tat, Kryz, and Tsakhur.

The Talysh comprise one of the largest linguistic minorities of Azerbaijan. There are about 350 Talysh villages, and in some areas, the Talysh consitute up to 95% of the population. Bilingualism in Azerbaijani is typical for the vast majority of the Talysh. According to Clifton, Deckinga, Lucht, & Tiessen (2003b), within Talysh settlements, Azerbaijani is used in formal situations where non-locals are present, while Talysh is used in informal situations for communication between locals. Older Talysh speakers are more likely to use Talysh in their everyday communication. Conversely, for younger people, Azerbaijani may become dominant. In ethnically mixed communities, only Azerbaijani is used in informal situations. There are significant differences between lowland villages with a stronger influence of Azerbaijani and mountain villages where Talysh is better preserved. Since the 1930s, Azerbaijani has been the language of instruction in Talysh schools. In 2003, Clifton, Deckinga, Lucht, & Tiessen (2003b) reported that a program of Talysh has been designed for grades 1 to 4 in homogeneous Talysh communities, and that there were plans to expand this program to include higher grades. Knowledge of Russian is widespread because many Talysh people go to Russia for temporary jobs. The use of Russian is significantly lower than the use of Talysh and Azerbaijani, particularly in the highland communities.

Tat has been displaced by Azerbaijani in many communities where it was traditionally spoken. In all Tat communities, average proficiency in Azerbaijani is high and Azerbaijani is the language of schools. According to another recent study by Clifton, Deckinga, Lucht, & Tiessen (2003a), the viability of the vernacular is tied to the economic viability of a given community and to the remoteness of the village. Tat may survive in large lowland

communities with ample resources, as well as in highland villages that are relatively stable economically. For example, in Qırmızı Qəsəbə, a settlement with a high concentration of Tats, the local language is still viable. It is the main language used by families, including in communicating with children, as well as between community members (Clifton, Deckinga, Lucht, & Tiessen, 2003a). The population as a whole also knows and uses Azerbaijani and Russian. In poorer villages, be they in highlands or lowlands, Azerbaijani has become the main language of Tat families, with parents communicating in it with their school-aged children. For the majority of Tat communities, Russian plays a secondary role; only a small portion of the population can speak Russian well.

Kryz (Nakh-Dagestanian, Lezgic) is spoken by a small group in northern Azerbaijan, including all age groups in the mountain villages of Hapıt, Əlik, Cek, and Qrız (Authier, 2009, p. 1). Outside Kryz villages (i.e., within families who live in lowland towns), the proficiency in Kryz varies from full fluency to passive understanding, and even to complete loss of the language. Clifton, Mak, Deckinga, Lucht, & Tiessen (2002) report fluency not only in Azerbaijani but also in Russian, though our own impressionistic observations in Alik show limited command of Russian among women and younger men (similar findings are reported for the Khinalug). Azerbaijani has always been the language of instruction in the highland villages of Hapıt, Əlik, Cek, and Qrız, and no Kryz classes have ever been offered in any of the villages.

Another Lezgic language, Tsakhur, is spoken in northern Azerbaijan. The Zakatala, Qax, and Sheki regions border southern Dagestan and were places of traditional seasonal migrations by the Tsakhur and Rutul peoples from Kurdul, Gelmets, Mikik, Kina, and some other villages. In the Qax region (briefly investigated by a field team led by two of us in July 2018), Tsakhurs hardly retain their language. Most Tsakhurs prefer Azerbaijani for communication, even within their family. Azerbaijani also dominates in those villages that are traditionally considered Tsakhur settlements (e.g., Kum). We do not have systematic data on Avar, Akhvakh, Rutul, or Lezgian spoken in Azerbaijan, but all these languages are reported to be strongly present in the northern areas bordering western Dagestan.

1.5.1.2   *Armenia*

Armenian is spoken by everyone in Armenia, except a minority with Russian as their first language. Western and Eastern Armenian are usually considered different varieties of Armenian or even different languages. With some exceptions (e.g., Hamshen or Kesab), Western Armenian dialects became extinct after the genocide of 1915. Most eastern dialects, on the other hand, are vigorously alive. These are considerably divergent varieties of Eastern Armenian, some of which are claimed to be unintelligible to the speakers of the standard language, especially the dialect of Artsakh (Nagorno-Karabakh). After the genocide, most refugees from Turkey fled to Europe and the United States, but some escaped to Armenia, which presumably led to an interesting and understudied situation of dialect mixing and leveling. Dialects are not taught at school. Modern Western Armenian is taught at certain schools, especially after the beginning of the civil war in Syria and the immigration of Syrian Armenians to Armenia. Classical Armenian is still used as the language of liturgy.

Besides Armenian, both active and passive command of Russian is widespread in Armenia. Certain generational biases are observed for L2 speakers, with younger generations being less proficient in Russian and more in English, and vice versa for older generations (Victoria Khurshudyan, personal communication, May 10, 2017 see also Dum-Tragut, 2013).

The presence of other ethnic groups in Armenia is limited (see Schulze & Schulze, 2016, for an overview). The largest minorities are Kurds and Yazidis, both bilingual in Armenian. Neo-Aramaic and Kurdish are taught at certain schools.

1.5.1.3 *Georgia*

In Georgia, Georgian is used in all domains of everyday life: education, work, life, and media. Standard Georgian is the language of instruction at schools and at universities. TV and radio broadcasts are in standard Georgian. Some TV channels have one- to two-hour informational blocks in Ossetic and Abkhaz. Megrelian and Svan are not used in the media or in schools.

There are some Russian schools with the full instruction cycle in Russian, although their number decreased considerably throughout the 1990s. English as a second language is gaining popularity but is still spoken by only a small part of the population. Georgians born before the late 1980s continue to use Russian as their main L2.

Even in border settlements, Georgians are usually monolingual, while some of their neighbors speak Georgian. Some Armenians and Azerbaijanis living in Georgia use Russian as a lingua franca. Dialects of Georgian are vigorous, not only in rural areas but also in towns.

There are several language minorities in Georgia, including speakers of the other Kartvelian languages (Megrelian, Svan, and Laz), as well as speakers of languages from other families, including Abkhaz, Azerbaijani, Russian, Ossetic, Kurdish, Armenian, Neo-Aramaic, and Batsbi.

The situation with Kartvelian minorities is complicated by the fact that these languages belong to the same family as Georgian and are sometimes considered Georgian dialects. By and large, their speakers identify themselves as Georgians (see Tuite, 2017b, p. 226, on Svan; Vamling, 2017, on Megrelian). In Georgia, all three languages remain mostly unwritten (see Tuite, 2017b, on attempts to write in Svan; see section 1.4 about attempts to write in Megrelian). In *Gal*, a newspaper published in Abkhazia, Megrelian is written in a modified Georgian script.

Svan is still spoken in its places of origin, Upper Svaneti and northwest Samegrelo, whereas Lower Svaneti and southeast Samegrelo are reported to have shifted to Georgian. In Mestia, according to recent reports, Svan is fluently spoken only by people older than 40. On the contrary, in the villages of Latali, Ipari, and Adishi, Svan remains the primary language of everyday communication, including for children (Kevin Tuite, personal communication, June 12, 2017). The loss of proficiency in Svan seems to be a relatively recent development. According to Nizharadze (1964, pp. 169–172, quoted by Hewitt, 1992), Georgian was primarily acquired by Svan men when they were working in the Georgian lowlands during winter. In 1964, out of the 290 men in the Svan village

of Ushguli, 160 knew Georgian; in K'ala, 199 of 219; in Ipari, 306 of 546. As early as 1870, the command of Georgian was reported for three to six men in each of the villages. According to Chirikba (2009, p. 28), Svans and Megrelians acquired Georgian in the 20th century largely due to the introduction of Soviet schools, where Georgian was the language of instruction.

Megrelian, also under strong pressure from Georgian, remains the dominant language of its area, at least outside institutional and administrative interactions (Alexander Rostovtsev-Popiel, personal communication, May 5, 2017). The situation of Laz is problematic in Sarpi, the only Laz village in southwestern Georgia, where locals are unwilling to admit they are Laz. In Turkey, Laz is much more viable. Few Turkish Laz, however, consider their L1 to be a social advantage, because it is unlikely to provide them a job and has low prestige in Turkey. Thus, young people tend to shift to Turkish.

Batsbi is a one-village language in the Georgian province of Kakheti. All speakers of Batsbi are bilingual in Georgian (Gippert, 2008). Similar to other Georgian minorities, the Batsbi prefer to be considered Georgians. According to the project of language documentation of the languages of Georgian minorities (http://dobes.mpi.nl/projects/svan/language), at the beginning of the 21st century only the generation older than 50 had perfect competence in the language; younger adults could understand it and were able to speak but did not use the language. There were no children who understood or used Batsbi.

Rural Armenian populations living in the Samtskhe-Javakheti district of southwestern Georgia are descendants of those Armenians who resettled there in the 19th century during the Russo-Turkish Wars. They are predominantly agriculturalists who tend to have a lower level of education in comparison to Armenians living in Tbilisi. Samtskhe-Javakheti Armenians are seldom proficient in Georgian and speak the Karin dialect of Armenian, also used in northern Armenia.

The Azerbaijani community in Tbilisi is very small. Georgian Azerbaijani populations are primarily rural agriculturalists concentrated in Kvemo Kartli, to the south of the capital near the border with Azerbaijan. Like the rural Armenians of Samtskhe-Javakheti, they are seldom fluent in Georgian and tend to have little formal education (Driscoll, Berglund, & Blauvelt, 2016).

1.5.1.4 *Abkhazia*

During the Soviet era and in the post-Soviet period, Abkhaz was taught at school and was the language of instruction until the fourth grade (i.e., up to the age of about ten). Starting from the fifth year of school, instruction switched to Russian (Chirikba, 2009, p. 7). Some subjects were taught in Abkhaz at the university level. Radio and broadcasting in Abkhazia are in Abkhaz and Russian.

L1 speakers of Abkhaz are the majority in Abkhazia. The largest minorities are Armenians and Megrelians. In the recent past, some Abkhaz people in the south of Abkhazia were bilingual in Megrelian. Bilingualism in Georgian, Armenian, Svan, or Greek is also reported, but is rare (Chirikba, 2009, p. 8). During the war with Georgia in 1992–1993, most Georgians, and many Megrelians and Russians, left the republic. Some

of them came back after the war, but in general the proportion of non-Abkhaz population in Abkhazia has declined. Local Armenians mostly speak Hamshen Armenian, but school instruction is in Standard Armenian in Armenian schools; Hamshen and standard Armenian are reported to be mutually unintelligible (Chirikba, 2008a).

Proficiency in Russian was very high during Soviet times and did not decrease significantly after the fall of the Soviet Union (Chirikba, 2009). Russian continues to be the main language of Abkhaz towns, especially among younger speakers.

1.5.2 Russia

In Ingushetia and Chechnya, Russian influence was stronger than in some other areas of the Caucasus because of the deportations, which were much more restricted and selective elsewhere (see chapter 2). As everywhere, Russification was stronger among urban populations. But while in the highlands of Dagestan, villagers continue to use their ethnic languages, this is not always the case in Ingushetia and Chechnya. Under Stalin, highland villagers were forcibly resettled, and the Chechen population decreased significantly. Those who came back in the 1950s often settled in the lowlands.

Chechnya is large both in terms of its area and its population. It is largely monoethnic, with very few Chamalal villages in the East, on the border with Dagestan, and equally few Kumyk and Noghay villages in the North. The command of Russian is very high and has not been affected even by the military conflicts with Russia. Russian is the main language of education and science, and the media are in both Chechen and Russian. Even in rural primary schools, the language of instruction is Russian; Chechen is present only in Chechen language and literature classes (Rus. *rodnoj jazyk i literatura*, 'native language and literature'; Zarina Molochieva, personal communication, June 12, 2017).

Formally, Ingush is the official language of Ingushetia, used in education, media, and literature. However, in the opinion of some experts, the language situation is unstable. Children in mixed families usually speak Russian. In the minds of many young people, the Ingush language and culture are associated with a rural, low-prestige lifestyle (Nichols, 2011). Nearly all Ingushes are bilingual in Russian. Many have a passive knowledge of Chechen. Active command of Chechen is infrequent and is more typical of mixed Ingush-Chechen families. Even less frequent is the command of Ossetic or Georgian which, in older days, used to be acquired by visiting marketplaces.

Chechen and Ingush are close, but there is no mutual intelligibility. Long periods of contact have resulted in passive bilingualism on both sides; oftentimes, people in a multilingual dialogue each speak their own language. If members of other ethnicities are present, Russian is used (Nichols, 2011).

The whole population of North Ossetia is bilingual in Russian. In towns, especially in Vladikavkaz, Russian is used more widely than Ossetic. Children born in Vladikavkaz usually do not speak Ossetic at all. Between 2001 and 2002, only 62 of 224 schools had primary school instruction in Ossetic. Only one of the 62 was located in Vladikavkaz (Kambolov, 2007, p. 22). Due to the local activist Tamerlan Kambolov, some

kindergartens have started offering classes where caregivers communicate with children exclusively in Ossetic.

There is a literary tradition, but according to Kambolov (2007), only about 10% of the population read books in Ossetic. Most prefer reading in Russian. The situation with radio and TV broadcasting is similar; although there are programs in Ossetic, Russian broadcasts are more diverse and are preferred by locals. Ossetic as taught at school is closer to the Iron dialect than to Digor. Newspapers, broadcasts, and theaters exist in both dialects (David Erschler, personal communication, June 2017).

The situation is similar in South Ossetia. Russian is still the primary L2, and many children in the urban areas do not acquire Ossetic. There are some Georgian schools, especially in the villages close to the Georgian border (Parastaev & Mearakishvili, 2016), but in most schools, instruction is in Russian.

The whole population of Adygea is bilingual in Russian. Adyghe remains the L1 for adults and children in villages, though code-mixing in Adyghe speech is becoming more and more frequent (Irina Bagirokova, personal communication, June 2017). The situation in Maykop, the capital of the republic, is that most children born there do not speak or understand Adyghe, even if their parents speak it as their L1. In kindergartens in both Maykop and the villages, caregivers speak Russian. At schools, a standard variety of Adyghe is used only in classes on the Adyghe language, literature, and customs. The standard variety is close to some dialects of Adyghe (Temirgoy and in some respects also Bzhedug), yet distant from other lects of the Adyghe-Kabardian continuum stretchings across the area where standard Adyghe is taught at school. Adyghe is sometimes used for academic purposes, including being one of the working languages at conferences in which representatives of the global Adyghe diaspora participate. There are radio and television broadcasts in Adyghe, as well as Adyghe theaters.

1.5.3 Dagestan

The Republic of Dagestan is special not only in terms of language density, extreme even on the scale of the Caucasus, but also because of the vitality of its minority languages. More than 40 languages are still spoken in the highlands. Endogamy is probably one of the factors which has sustained linguistic diversity. Traditionally, people from highland rural communities preferred to find marriage partners in the same village; often, there existed strict prohibitions on marrying out (Comrie, 2008). The pattern was probably not ubiquitous; according to Wixman (1980, p. 94), it was not strict in Tabasaran communities, where mixed marriages with Lezgians were not uncommon. Our own field experience revealed strict endogamy in Central and Northern Dagestan, and less strict endogamy in Southern Dagestan. For example, in the Rutul village of Kina, many men married women from a neighboring Tsakhur village; there is extensive intermarriage between the Agul village of Khpyuk and the Lezgian village of Ursun, though we cannot be sure that these are not innovative developments. In any

case, marrying to another village was only possible for women; men did not normally move to the village of their wives.

A large-scale sociolinguistic survey started in 2011 made available a massive amount of data on multilingualism in Dagestan, both in its present state and its recent past (see http://multidagestan.com). Such data are unavailable for other areas in the Caucasus. There is a huge difference in the patterns of interethnic communication in Dagestan before and after the advent of Russian in the middle of the 20th century. Prior to the Soviet period and the spread of Russian, the Caucasus had never had one single lingua franca (Chirikba, 2008b, p. 30). The language of interaction had always been negotiated between particular communities.

The most frequent pattern of linguistic interaction between neighboring villages used to be "neighbor multilingualism," where one of the languages of two neighboring villages would be used. If residents of two villages located within walking distance spoke different first languages, multilingualism was usually asymmetrical, meaning less than half of the population of one village could speak the language of their neighbors, while more than half of their neighbors were bilingual (Dobrushina, 2011a, 2013). This was clearly the case of Mehweb: 97% of the Mehweb-speaking population born before 1919 spoke Avar, whereas in the neighboring Avar village, Obokh, only 8% spoke Mehweb.[7]

In some parts of Dagestan, mainly a lingua franca was used between villages with different languages. This was the case in southern Dagestan, where the communities near the border with Azerbaijan were proficient in Azerbaijani. For example, the residents of the Rutul village Kina communicated with their Tsakhur neighbors from Gelmets in Azerbaijani. Another lingua franca was Kumyk, the Turkic language of the lowlands. Residents of the villages speaking Kadar Dargwa (e.g., Chabanmakhi, Vanashimakhi, Chankurbe) used Kumyk to speak with their Avar neighbors from the village of Durangi. In some parts of the central and northern highlands, Avar was the lingua franca. A striking example is the use of Avar in the Akhvakh district, where speakers of several Andic languages, Karata, Akhvakh, Bagvalal, and Tukita, much more closely related to each other than to Avar, communicated exclusively in Avar and rarely spoke each other's language (Dobrushina & Zakirova, 2019).

The choice of dominant language in asymmetrical bilingualism or as a general lingua franca was governed by many factors, including the language of the local marketplace, population size in the area and also, more visibly, by the relative altitude of the settlements. That the altitude of a language community may correlate with asymmetrical bilingualism is known from the patterns of the vertical hierarchy of Andean cultures (Murra, 1968). This was independently suggested in Caucasian studies (Lavrov, 1953; Volkova, 1974; see also Nichols, 2004). According to Nichols (2013, p. 43), "Highlanders learned lowland languages for market communication and also because a good portion of the adult working-age male population was transhumant, spending

---

[7] All numbers in this section come from the survey within the project *Atlas of Multilingualism of Dagestan* (http://multidagestan.com).

the winter months in lowland pastures; but lowlanders had no need to learn highland languages."

Apart from competence in neighboring languages, people in a number of areas in Dagestan also spoke some more distant languages (i.e., those which were spoken as L1 beyond their immediate neighborhood). Such languages were acquired due to transhumant (nomadic) husbandry, seasonal jobs or services, or trade. For example, in Archib (Archi), Shalib (Lak), and Chittab (Avar), people used to go to Azerbaijan to work as tinners. As a result, in these villages about 30% of the population born before 1919 could speak some Azerbaijani. In Balkhar, a Lak village, about 15% of the population born before 1919 could speak Avar, because Avar-speaking areas were important for selling pottery, for which Balkhar was famous.

The traditional patterns of multilingualism differed across genders. Languages of distant communities were spoken almost exclusively by men, because the practices which led to the acquisition of distant languages (such as tinning) were only typical of men (Dobrushina, Kozhukhar, & Moroz, 2019). Communication with immediate neighbors was frequent among men and women alike, so the gender bias was much less pronounced in the knowledge of geographically neighboring languages. Figure 1.1 shows that the percentage of those who spoke Avar in the Andi village of Rikvani was roughly the same among men and women. However, the gender bias is much more apparent for distant languages such as Russian, Chechen, and Kumyk.

Before the Soviet period, Dagestan had almost no secular schools. Some villages had madrasahs, where Arabic literacy was taught. Basic knowledge of the Quran and Classical Arabic script was also obtained through personal training provided by educated people, often the mollahs. As mentioned in section 1.4, Avar, Lak, Dargwa, Lezgian, and Kumyk used the Arabic script for writing. There are also some documents
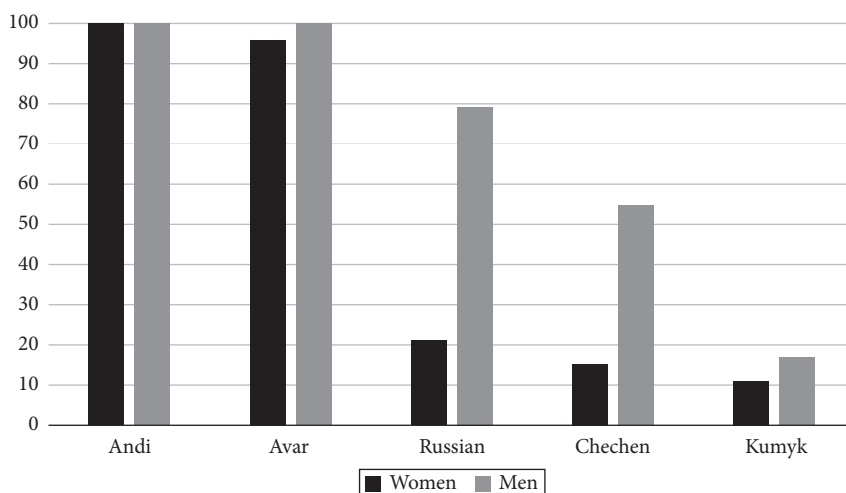

FIGURE 1.1. Languages spoken by men and women born between 1889 and 1940 in Rikvani (Andi). Data collected in interviews in Rikvani in 2015.

in minority languages written with the Arabic script. According to our research, in the generations born between 1880 and 1920, between 20% (Kina, Darvag) and 50% (Karata, Archib) could read the Arabic script (see http://multidagestan.com; also Bennigsen & Lemercier-Quelquejay, 1985, on Arabic in Dagestan).

Knowledge of the Arabic script did not imply command of Classical Arabic as a language. Many people were able to read Arabic letters but could not understand the meaning of the text. This is also true today—some Dagestanians can read the Quran, but knowledge of Arabic is rare.

According to historical evidence, Russian was spoken by less than 1% of the Dagestanian population in the end of the 19th century (Svod, 1893; Volkova, 1974, p. 31). In the 1930s, Soviet schools were opened in most Dagestanian villages. Russian was one of the main subjects and the primary language of instruction. The curriculum was compulsory. Most people thus acquired Russian and started reading and writing in Cyrillic. Figure 1.2 shows command of Russian in four villages of Dagestan as a function of the year of birth. In the generations born after 1950, almost everyone speaks Russian. The expansion of Russian has strongly enhanced the level of literacy and has severely endangered the patterns of local multilingualism.

From the middle of the 20th century, Russian has become the language of higher education and upward social mobility in Dagestan. Younger generations speak Russian instead of the languages of their neighbors. Figure 1.3 shows the dynamics of four languages in the village of Mehweb as a function of the date of birth. The native language of the village, Mehweb Dargwa, is spoken by all residents. Neighboring languages, Lak and Avar, started their decline in the 1930s and 1980s, respectively. Avar persisted longer because it was taught at school, together with Russian. All villagers born after 1950 speak Russian.
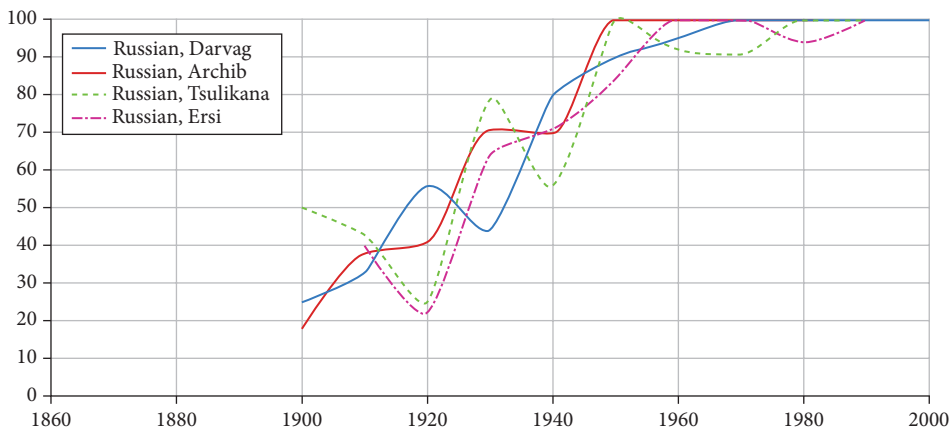

FIGURE 1.2. Command of Russian among people born between 1900 and 2000 in Darvag, Archib, Tsulikana, and Ersi (generated via http://multidagestan.com).
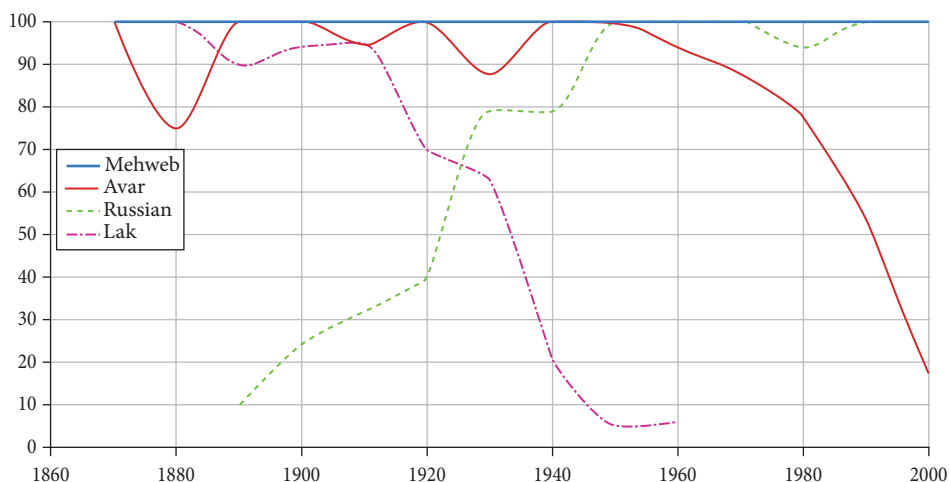
FIGURE 1.3. Languages spoken in Mehweb among generations born between 1870 and 2000 (generated via http://multidagestan.com).

At present, Russian is the main language of interethnic communication in many highland areas and in the vast majority of lowland settlements. The latter are meeting places for people with different linguistic repertoires. An important factor in the spread of Russian was that townspeople quickly abandoned traditional prohibitions on inter-ethnic marriages. Mixed marriages are still much more frequent among urban dwellers than in rural populations. The villagers who marry women from other ethnicities (whom they often meet in the lowlands, for example, during their graduate studies) move to towns. Such mixed families choose Russian as the main means of family communication. Under the influence of local languages, and probably also because of the drastic decrease of the ethnically Russian population in the 1990s, Russian used by Dagestanians has acquired special phonetic, morphological, and syntactic features and can be considered a distinct ethnolect (Daniel, Dobrushina, & Knyazev, 2010).

At present, migration is the main social process which has a strong impact on the loss of local languages. Migration to the lowlands was partially initiated by the Soviet authorities as an economically driven policy to relocate highlanders. From the 1950s through the 1970s, the residents of some mountain villages were put under pressure to move down and were financially supported to build their new houses in the lowlands (Karpov & Kapustina, 2011). As a result, many villages in various areas of Dagestan are now underpopulated or deserted. For instance, the speakers of Hinuq, a Tsezic minority language, started moving to the village of Monastirsky in the Kizlyar district of Dagestan in 1986 (Forker, 2013b).

Quite frequently, migration to the lowlands was linked to seasonal herding of village sheep in lowland pastures far away from the home village, so-called *kutans*. Originally, kutans were only used in winter. In other seasons sheep were pastured in the mountains. Nowadays, people often prefer to stay in lowland settlements for the whole year, thereby

establishing new villages. This process was boosted by the economic collapse of agriculture in the post-Soviet period (Kazenin, 2012).

The ethnic and linguistic composition of these new villages varies significantly. Sometimes they host people from different parts of Dagestan with different home languages. For example, the village of Druzhba, founded by Dargwa resettlers in the Kayakentsky District, also hosts Tabasarans, Aguls, Laks, and a few Lezgians, Russians, Kumyks, and Avars. According to Rasul Mutalov (personal communication, May 10, 2017), people use Russian for interethnic communication and practice mixed marriages. The situation in Druzhba inevitably leads to the loss of local languages. Some new lowland settlements maintain ethnic homogeneity which was typical of traditional Dagestan and consolidate residents of several villages who speak different dialects of the same language. Personal communication with residents suggests that they are aware of dialect levelling in these villages, but at least the local language is not entirely lost. Neither language loss nor dialect leveling in the lowlands has been studied; one study of code-switching in Sanzhi-Itsari spoken in Druzhba is a notable exception (Forker, 2019b).

1.6 Language Contact in the Caucasus

Caucasian languages share a number of common properties (see also the Introduction to this volume), for example, ergative alignment and the presence of ejectives. It is true that the three endemic families are overwhelmingly ergative. Ejectives seem even more significant in that they have spread to non-endemic families—such as Ossetic (Iranian)—and are argued to be present in some dialects of Kumyk, Karachay-Balkar, and Azerbaijani (Turkic; see Fallon, 1998, p. 320).

Yet, shared linguistic features of the Caucasian languages as a whole are less striking than, for instance, shared features in the Mesoamerican linguistic area (see Campbell, Kaufman, & Smith-Stark, 1986). Tuite (1999) argues that the "three ergativities" in the Caucasus are morphosyntactically very different, each more similar to ergative patterns attested elsewhere in the world than to one another (see also the Introduction to this volume). Ejectives may be acoustically different even in languages of the same family (see Grawunder, 2017, for a thorough overview of the Caucasus as a phonetic contact area).

It seems that, as with many other areas of language contact, it is more useful to look for features shared by some, rather than all, or even most, languages. It also seems that the boundaries of language families in the Caucasus are less transparent to language contact than in some other areas, such as Mesoamerica or Amazonia. Strong cross-family contact is observed in Ossetic (influence from Nakh languages) and in Armenian (influence from Iranian and Turkic languages). Nakh-Dagestanian shows lexical contact in border languages, such as lexical borrowings from Georgian to Tsezic, while there are many features that seem to result from structural alignment within the Nakh-Dagestanian family itself.

1.6.1 Lexical Borrowing

Loanwords are an obvious point of interest of historical studies on Caucasian languages. Lexical contact in Armenian and Georgian is a topic of interest for traditional comparative linguists (for a recent overview of Armenian etymologies, see Martirosyan, 2010; for loanwords in Nakh-Dagestanian and Northwest Caucasian, see Höhlig, 1997; Khalilov & Comrie, 2010; Klimov & Khalilov, 2003; Nikolaev & Starostin, 1994; Shagirov, 1989; for Kartvelian etymologies, see Klimov, 1998a; various Caucasian etymologies are summarized at http://starling.rinet.ru). Lexical borrowings in the Caucasus are included in the World Loanwords Database (WoLD) project whose perspective is mainly sociolinguistic (Haspelmath & Tadmor, 2009). In addition to several languages with a long tradition of comparative lexical research, such as Ossetic and Armenian (Indo-European), Georgian (Kartvelian), or Kumyk and Azerbaijani (Turkic), WoLD also covered less well-studied and/or minority languages, Archi and Bezhta (Nakh-Dagestanian; see Chumakina, 2009a, 2009b; Comrie & Khalilov, 2009a, 2009b). Taking Nakh-Dagestanian languages as an example, based on WoLD and other sources, the case may be made for distinguishing between "vertical borrowing," with loanwords coming from "external," culturally dominant languages, such as Arabic, Turkic, Iranian, and, more recently, Russian, and "horizontal borrowing," with loanwords between branches of the same family, or, more rarely, across family borders. Horizontal borrowing may occur from locally important languages to minority languages, such as Lak borrowings in Archi, or Avar and Georgian borrowings in Bezhta (cf. Bezhta ɣadri 'embers' (Georgian ɣadari), kanłi 'light' (Avar kanłi) versus ulka 'country' (Turkic ülke), insan 'someone, person' (Arabic insa:n), picka 'match' (Russian spíčka)) (Comrie & Khalilov, 2009b). Vertical borrowings may have entered minority languages via locally important languages (Avar in the case of Bezhta), which blurs the sociolinguistic relevance of this difference.

   An ongoing project on quantificative analysis of horizontal lexical borrowings in Dagestan has already yielded some results (see Chechuro, 2018). In the south, where bilingualism in Azerbaijani was almost universal (including some villages shifting to Azerbaijani, see section 1.5), the presence of Turkic loanwords is by far more visible. Across Andic languages, bilingualism in Azerbaijani or Kumyk was much less widespread, sometimes even non-existent, and Turkic loans may have been acquired via contact with other Nakh-Dagestanian languages. Andic languages show very strong influence from Avar, which served as the local lingua franca (see Magomaeva & Khalilov, 2005; Dobrushina & Zakirova, 2019). The results of this ongoing project correlate with the data on local multilingualism (http://multidaghestan.com).

   During fieldwork in 2018, we collected lexical lists in the Qax region of Azerbaijan and discovered that the Azerbaijani dialect of the villagers of Ilisu contained quite a few lexical items with ejectives (previously reported in Aslanov, 1974, referring back to a 1945 field report by Shiraliev).[8] The presence of these items was registered in interviews with the villagers of Ilisu or with people born there. All these items were nouns designating

---

[8] Unfortunately, the latter source was unavailable to us.

objects such as utensils or plants; however, data collection was not systematic across the lexicon, so borrowed items from other lexical categories may also exist. Almost all these words have been identified as Tsakhur lexical items that are known to the Tsakhur in Dagestan. This is unsurprising, because Ilisu is known to be an old center of Tsakhur presence in the territory of present-day Azerbaijan (the former Sultanate of Elisu). Surprisingly, none of the interviewees identified these lexical items as Tsakhur loans; they were all claimed to be merely characteristic of the local dialect of Azerbaijani.

Additionally, in their speech, voiced consonants in some Azerbaijani words of clearly Turkic origins had a more or less articulated ejective realization (such as *t'ana* for Azerbaijani *dana* 'calf', used almost consistently). In this case, the usual direction of contact-induced change (from Turkic to Nakh-Dagestanian) is inverted, most probably due to the fact that it resulted not from bilingualism in a dominant language but from language shift from a minority language (also inferred by Aslanov, 1974). Reconstructing local patterns of language shift may thus provide insights not only in local ethnic history but also in theories of phonetic contact.

1.6.2 Contact-Induced Change in Grammar: An Overview

Outcomes of lexical contact in the Caucasus, such as various types of pattern replication in lexical domain, have not yet been studied on a systematic basis.[9] Similarly, pattern replication in grammar has only attracted considerable attention in recent decades, and the studies have so far been rather selective. Donabédian (2000, 2018) provides an overview of contact phenomena in the history of Armenian grammar, including its evolution from Classical Armenian into the drastically different system of the modern language. For a discussion of contact phenomena in Northwest Caucasian, see Höhlig (1997). There is nascent research on contact phenomena in Nakh-Dagestanian, with its multitude of languages and contact situations inevitably leading to structural convergence, but many languages of the area have not been documented thoroughly enough to identify contact-induced changes. Some recent studies include Dobrushina (2017), who argues that the origins of a specific use of volitional moods in subordination are due to contact with Turkic; Authier (2010) who considers contact-induced morphology in Kryz (both pattern and matter copy); or the claim that the evolution of person in Batsbi is due to strong influence from Georgian (Kojima, 2019). For other studies, see also Belyaev, 2019; Desheriev, 1953, p. 85; Maisak, 2019a, 2019b. Maisak (2016b) is an overview of contact-induced change in Nakh-Dagestanian (as well as the pertinent methodological issues), and the discussion of contact-induced change in Nakh-Dagestanian in this section largely follows his overview.

Given the high degree of neighbor multilingualism (see section 1.5), contact-induced phenomena in grammar are expected. However, shared grammatical features which arise via the borrowing of abstract grammatical patterns are notoriously difficult to

---

[9] See Haspelmath (2009) and Koptjevskaja-Tamm & Liljegren (2017) for a general discussion.

attribute to language contact. In selecting examples of apparent contact-induced phenomena, we chose such phenomena where shared features cannot be easily interpreted as inherited from the protolanguage—even if shared by languages that have never been in direct contact. These must have spread through chains of local bilingualism; a feature is borrowed from one language to a neighboring language, and from that latter language to yet another language. Sections 1.6.3 and 1.6.4 give examples of such features.

1.6.3 Alternations in Recipient Marking

Nakh-Dagestanian languages are cross-linguistically rare in having two distinct patterns of recipient encoding, which we will refer to as the "dative strategy" (with *ez* 'to me' in (1)) and "lative strategy" (with *ju-w oq'er-mu-ra-k* 'to this pauper' in (2)). Consider the constructions in Archi in (1) and (2):

(1)   Dative Strategy in Archi
    jella   *wiš*   *ja-r*   *laha*   *pari*   *χanum-li,*
    thus   you.PL.GEN   this-2   girl.OBL(ERG)   P.   X.-OBL(ERG)
    at'u-li,   *jeb*   *sot:-or*   *LO,*   *χir*
    NPL.cut.PFV-CVB   this.PL   bead-PL   NPL.give.PFV   behind
    da-qˤa-li,   *ez*   *LO*
    2-come.PFV-EVID   NPL.I.DAT   NPL.give.PFV
    'Thus, this girl of yours, Pari Hanum, tore off her necklace, ran after me and gave it to me.' (Archi Electronic Corpus)

(2)   Lative Strategy in Archi
    k'an   *harak-du-t*   *iq-n-a*   *ja-r*   *ɬ:anna*
    most   before-ATTR-4   day-OBL-IN   this-2   woman.OBL(ERG)
    č'ut   *bo-LO-li*   *ju-w*   *oq'er-mu-ra-k*   *da‹b›χi-s*
    jug   3-give.PFV-EVID   this-1   pauper-OBL.1-CONT-LAT   ‹3›hit-INF
    'On the (very) first day this woman gave this pauper a jug (of butter) to churn.' (Archi Electronic Corpus)

At first glance, the difference is between permanent transfer to a recipient ("dative strategy") and temporary transfer ("lative strategy"). The event in (1) involves the delivery of a gift, while (2) describes temporary transfer. Upon closer scrutiny, the distinction is more complex. The lative strategy is also used when the object is given back to its owner (*šahʁuli ap:p:as-a-l-di* 'to Shakhguli Abbas' in (3)) or when it is given to a final recipient by a mediator (*d-is ɬ:anna-ra-k* 'my wife' in (4)).[10]

---

[10] For further details on this phenomenon, see Daniel (2019); Daniel, Khalilova, and Molochieva (2010).

(3) Lative Strategy in Agul

 me *ruš* *š-u-ne* *fac-u-na* *qa-i-ne*
 this girl go-PFV-AOR catch-PFV-CVB RE-give.PFV-AOR
 šahʁuli *ap:p:as-a-l-di* *ħajwan.*
 Shakhgul Abbas-OBL-SUPER-LAT horse
 'This girl went there, caught the horse and gave it back to Shakhguli Abbas. (Agul Electronic Corpus)

(4) Lative Strategy in Archi

 un *daki* *LO-t'u* *d-is*
 2SG(ERG) why 4.give.PFV-NEG 2-I.GEN
 ɬ:anna-ra-k *is* *amanat* *bo-li*
 woman.OBL-CONT-LAT 4.I.GEN present say.PFV-EVID
 'Why didn't you give my wife the present that I gave (to you for her)—he asked.' (Archi Electronic Corpus)

These subtle semantic distinctions are consistent across all languages of the family, from the Nakh languages in the northwest to the Lezgic languages in the north of Azerbaijan. The only known Nakh-Dagestanian language that lacks this distinction is Udi, a Lezgic outlier in central Azerbaijan; Udi was cut off from the geographic continuity of Nakh-Dagestanian languages for centuries. From Nakh languages, the distinction also expanded into both Iron and Digor Ossetic (Belyaev, 2019; Belyaev & Daniel, 2014a, 2014b). On the other hand, this feature is probably not inherited from an ancestral language, because the lative strategy in different languages is encoded by different morphological means that are certainly not cognate in all languages. Even in the two Lezgic languages considered here, Archi and Agul, the spatial markers used in the lative strategy are semantically different (CONT vs. SUPER), and neither of the lative markers is cognate.

Marking the distinction between the two types of transfer is cross-linguistically rare. Outside the Nakh-Dagestanian family, a similar pattern is known to exist in Dravidian.

1.6.4 Ordinal Numerals Formed with 'Say'

In Mehweb Dargwa, ordinal numerals are formed by combining the root of the cardinal numeral with the attributive form of the verb 'say' (lit. 'one (to which one should) say N'):

(5) Mehweb Dargwa
 k'ʷi-e-s-i
 two-say.IPFV-INF-ATTR
 'second'

This way of forming ordinal numerals is attested in Lezgic, Dargwa, Lak, and Tsezic, as well as in Akhvakh and some dialects of Avar—that is, all branches of the family except

Nakh and Khinalug (in Khinalug, Azerbaijani loans are used). Thus it seems that the construction is common to the languages from central to southern Dagestan (Maisak, 2016b). As further argued in a comprehensive family-wide overview of the data in Nasledskova and Netkachev (to appear), this strategy of forming ordinal numerals does not seem to be cross-linguistically frequent (if attested at all). The verb 'say' used in this construction is not necessarily cognate in all branches and comes in slightly different morphological forms. This suggests another case of structural alignment which could have spread through chain bilingualism.

1.6.5 'Find' as an Epistemic Auxiliary

In Mehweb Dargwa, the verb 'find' is used in several constructions related to epistemic and evidential domains.

(6)   Direct Evidential Use of 'Find' in Mehweb Dargwa
    šejťat-une-jni   *id-di*   *d-arʔ-i-le*
    devil-PL-ERG   this-PL   NPL-gather.PFV-PST-CVB
    ar-d-uχ-i-le   *d-arg-i-le*   *le-r*
    PV-NPL-bring.PFV-PST-CVB   NPL-find.PFV-PST-CVB   COP-NPL
    '(When he came back to the place where he dropped the gold) he discovered that the devils picked it up and carried it away.'

(7)   Epistemic Use of 'Find' in Mehweb Dargwa
    abzul=la   *huj-ni*   *abx-i-le*   *d-urg-a-re*   *ʔʷaʕnd*
    all=ADD   road-ERG   open-PST-CVB   NPL-find.IPFV-IRR-PST   PTCL
    b-ikʼ-a-re   *b-emž-ul-le.*
    HPL-come.IPFV-POT-PST   HPL-be.hot.PFV-PTCP-CVB
    '(The windows) must have been open all the way (lit. on all the road), otherwise we would have suffocated in the heat.'

(8)   Conditional Use of 'Find' in Mehweb Dargwa
    nu   *qʼoʕj-ħe*   *w-arg-a-ǩa,*   *uk-iša.*
    I   go.IPFV.PTCP-IN(LAT)   M-find.PFV-IRR-COND   M.lead.PFV-FUT.EGO
    'If (you go) where I go, I will give you a ride.'

In these three cases, three different forms of the verb 'find' are used as quasi-auxiliaries with lexical verbs to convey three very different meanings. The perfect form in (6) is used to convey direct evidentiality; in (7), the future form is used for presumptive evidentiality (i.e., deduction by knowledge of the real world); and in (8), the conditional converb is used in a periphrastic conditional. All three constuctions with 'find' are also attested in some Andic and Tsezic languages, Avar, and Archi. Avar, Andic, and probably Tsezic are distantly related within the family and might have inherited the construction. Archi and Mehweb Dargwa belong to other branches of Nakh-Dagestanian, but

form exclaves in Avar-speaking territories. They have probably been in a state of constant bilingualism in Avar for many centuries, so the presence of the 'find' constructions makes a strong case for contact-induced change. The presence (or absence) of these uses of 'find' in other Nakh-Dagestanian languages still needs to be checked, but it does not seem to be attested in Lezgic languages other than Archi.

1.6.6 Differential Object Marking in Udi

Udi, a Lezgic outlier of central Azerbaijan, features differential object marking (DOM). In (9) the form is *ereq:-a* 'hazelnut (Dat)', and in (10), it is *ereq:* 'hazelnut':

(9)   Differential Object Marking in Udi
       (…) *ajiz-e*   *60-ǯi usen-χo-st:a*   *ereq:-a*   *üše*
          village-LOC   60-ORD year-PL-AD   hazelnut-DAT   at.night
       bašq:-esun   modaχ=e=j.
       steal-NMLZ   common=3SG=PST
       'In the village, it was common in the sixties to steal hazelnuts in the night.'
       (Kasyanova, 2017, p. 630)

(10) Differential Object Marking in Udi
       hälä   ereq: toj-st:-a   bar=te=t:un=ne=j
       yet   hazelnut   sell-LV.INF-DAT   let=NEG=3PL=LV.PRS=PST
       'It was not permitted to sell hazelnuts.' (Kasyanova, 2017, p. 630)

Kasyanova (2017) argues that the primary factor triggering DOM in Udi is the referentiality of the object. Definite objects tend to be marked with a dative, while other objects are in the nominative (unmarked). Kasyanova notes that DOM is not attested in other Nakh-Dagestanian languages, but it is found in Azerbaijani, Farsi, and Armenian. DOM, while widespread cross-linguistically, seems to be rare in ergative languages (Malchukov, 2006). Thus, it is very likely that its presence in Udi is due to contact with surrounding languages. In Turkic languages, DOM is also sensitive to the referentiality of the object. In Armenian, the primary condition on DOM is animacy, but within animate nouns it is highly referential nouns that are marked. Whatever the exact scenario, the emergence of differential object marking in Udi seems to be a relatively clear case of a contact-induced phenomenon.

1.7 CONCLUSION

On the verge of the 21st century, "the amazing historical stability of the linguistic landscape of the Caucasus is coming to an end" (Gippert, 2008, p. 161). The populations in Azerbaijan, Georgia, and Armenia are rapidly losing proficiency in Russian.

The respective national languages of these countries have been bolstered by recent transitions to independence. They are now used in most domains of education and culture. At the same time, minority languages of Azerbaijan, Georgia, and Armenia did not benefit from independence. The dominant language simply switched from Russian to the national language of the country: Azerbaijani, Georgian, and Armenian, respectively. Most minority languages are not taught at school, are not used for reading or writing, and have low prestige in their communities. In Abkhazia and Ossetia, Russian has stayed strong, and the use of Abkhaz and Ossetic has not expanded into new domains.

The parts of the Caucasus which remained part of Russia are all characterized by the prevalence of Russian. Local languages (e.g., Ingush, Chechen, Ossetic, Adyghe, and many languages of Dagestan) are well-preserved in rural areas but are largely lost in cities. Since the majority of the population of the Caucasus live in villages, linguistic diversity is still preserved, but in some areas, local languages are endangered, primarily because of migration to the lowlands.

Acknowledgements

···

NORTH CAUCASUS
Regions and Their Demography

···

KONSTANTIN KAZENIN

THE goal of this chapter is to provide basic information on the ethnic composition of the North Caucasus, with emphasis on changes occurring over the last several decades. The chapter compares census data from different Soviet and post-Soviet years and discusses the major reasons which underlie population shifts (such as voluntary and forced migrations and changes in birth rate). Major interethnic conflicts of recent decades in the North Caucasus are also surveyed. Section 2.1 discusses northeastern regions; section 2.2 discusses the northwestern ones. Some of the discussion in this chapter complements the material in chapter 1.

2.1 NORTHEASTERN CAUCASUS

The northeastern Caucasus includes three republics within the Russian Federation: Dagestan, Chechnya, and Ingushetia. All three have a small Russian population (less than 5%) and a very high proportion of Sunni Muslims (more than 90%). However, they differ considerably in ethnic composition and recent patterns of migration.

2.1.1 Dagestan

Dagestan is the most populous and diverse region of the North Caucasus. Its population was 3,041,900 in 2017, and its ethnic diversity is the richest in Russia (the exact number of peoples depends on how certain minorities are grouped). Dagestan is the easternmost republic in the North Caucasus, bordered by the Caspian Sea to the east and connected by road to Azerbaijan in the south. Mountains cover almost half of

Dagestan, and population density in the mountains is much higher than in other North Caucasian republics.

Table 2.1 shows the ethnic composition of Dagestan according to the Soviet censuses of 1959 and 1989 and the Russian census of 2010 (some minorities are not included). Six groups account for nearly 85% of the entire population: Avars, Dargwa, Kumyks, Lezgians, Laks, and Tabasarans. Among them, Avars, Dargwa, Laks, and Tabasarans originate from local mountain ranges and have migrated in great numbers to the valleys since the second half of the twentieth century. Lezgians are the most numerous people of southern Dagestan, residing in both mountains and valleys. They also form a majority in northeastern Azerbaijan. Kumyks reside almost exclusively in the lowlands of the republic and in the foothills of the Caucasus Mountains. As Table 2.1 shows, all six major ethnic groups of Dagestan experienced at least 200% growth between 1959 and 2010. That growth was mainly due to a considerable decrease in mortality, especially infant mortality, in the 1950s and 1960s (Muduev, 2003, p. 41), after which the birth rate remained high for several decades (Kazenin & Kozlov, 2016). A radical decrease in birth rate starting from the 1990s suggests that the net population increase observed from 1959 to 2010 will not continue in the near future.

Minorities originating from the mountains underwent a similar increase in population between 1959 and 2010. This includes the Aguls, Rutuls, and Tsakhurs in the highlands of southern Dagestan. Table 2.1 does not include the Andic groups or the Tsez,

Table 2.1 Major Ethnic Groups in Dagestan in 1959, 1989, 2010

	1959	1989	2010
Avar	239,373 (22.5%)	496,077 (27.5%)	850,011 (29.2%)
Dargwa	148,193 (13.9%)	280,431 (15.6%)	490,384 (16.9%)
Kumyk	120,859 (11.4%)	231,805 (12.8%)	431,736 (14.8%)
Lezgian	108,615 (10.2%)	240,370 (11.3%)	385,240 (13.2%)
Lak	53,451 (5.0%)	91,682 (5.1%)	161,276 (5.5%)
Tabasaran	33,545 (3.2%)	78,196 (4.3%)	118,848 (4.1%)
Agul	6,378 (0.6%)	13,791 (0.8%)	28,054 (1.0%)
Rutul	6,566 (0.6%)	14,955 (0.8%)	27,849 (1.0%)
Tsakhur	4,278 (0.4%)	5,194 (0.3%)	9,771 (0.3%)
Chechen	12,798 (1.2%)	57,877 (3.2%)	93,658 (3.2%)
Noghay	14,939 (1.4%)	28,294 (1.6%)	40,407 (1.4%)
Russian	213,754 (20.1%)	165,904 (9.2%)	104,020 (3.6%)
Azeri	38,224 (3.6%)	75,463 (4.2%)	130,919 (4.5%)
Tat	19,155 (1.8%)	25,978 (1.4%)	455 (0.1%)

Source: Soviet population censuses of 1959 and 1989, Russian population census of 2010.

ethnicities closely related to the Avars who reside in the western Dagestanian moun-
tains. They were not even registered as separate ethnicities when the Soviet censuses
started in the area in 1939. In the censuses held after the collapse of the Soviet Union,
some registered under separate ethnonyms, and others preferred to keep their Avar
identity (see Kazenin, 2002a; on the expansion of the list of ethnicities in post-Soviet
censuses, see Bogojavlenski, 2008).

Dagestanian Chechens reside mainly in and around the town of Khasavyurt, near the
border of Chechnya. After the exile to Central Asia and Kazakhstan imposed upon them
by Soviet authorities from 1944 to 1957, Chechens were not allowed to return to some of
their rural homelands in Dagestan, instead settling in neighboring areas. The Auch
municipal district of Dagestan, which was, for the most part, inhabited by Chechens in
1944, has yet to be restored. This has caused interethnic tensions in those areas histori-
cally inhabited by Chechens (Kazenin, 2013b).

Official statistics classify two distinct groups of people as "Azeri." One group consists
of Azeris mainly residing in the old quarters of Derbent, the biggest town in southern
Dagestan, located on the Caspian Sea. This population is composed mostly of Shia
Muslims. The second group, the Sunni Azeris, reside in rural valleys in southern
Dagestan and in the towns of Derbent and Dagestanskiye Ogni. The two groups differ in
their dialects and in self-identification.

The Tat population in Dagestan has shown the most drastic decrease in recent
decades (from over 25,000 in 1989 to less than 500 in 2010). Of the three confessional
groups (Jewish, Muslim, and Christian) associated with this Iranian minority, mainly
the Jewish Tats are present in Dagestan. Dagestanian Tats are mainly concentrated in the
town of Derbent; after the collapse of the Soviet Union, many moved to Israel or scat-
tered throughout regions of Russia.

Russians in Dagestan have settled primarily in two areas. One includes major cities—
mainly the capital, Makhachkala, and the capital's satellite, Kaspiysk. A boom in the
Russian population took place during the Soviet era, when officials supported migration
into the region by industrial workers and engineers who were employed at the plants
built there. Russians are also concentrated in the north of the republic, mainly in the
town of Kizlyar and the rural valleys around it. Russian Cossacks have inhabited that
area since at least the eighteenth century. Intensive migration of Russians out of the
region started in the 1960s and 1970s and resurged after the collapse of the Soviet Union
(Belozerov, 2005, pp. 54–108). Ongoing movements of indigenous peoples from the
mountains into areas where Russians used to live is the most commonly accepted reason
for the migration of Russians out of the region (see more in section 2.1.2), in addition to
cuts in the number of workers at those plants where many Russians were employed.

Noghays are a Turkic people found in Dagestan and some other areas of southern
Russia. In Dagestan, the largest group of Noghays lives in the northernmost valleys near
the border with Stavropol. There are also Noghays living in some areas close to
Makhachkala, along the Sulak River by the Caspian Sea. The Noghay population's modest
growth between 1989 and 2010—compared to that of the other major peoples in
Dagestan—can be explained by an earlier decrease in the birth rate, as well as their intense

migration out of Dagestan (mainly to western Siberia, but also to the Astrakhan region in the southern part of the Volga River Basin).

Starting in the 1950s, indigenous groups have moved in great numbers from the mountains (highlands) to the valleys (lowlands). This migration and the urbanization of local populations have radically changed the ethnic composition of the Dagestanian lowlands. Three major waves of migration during the Soviet era can be identified, each differing in their causes and conditions (see Osmanov, 2000; for a general economic and demographic overview of post-Soviet Dagestan, see Radvanyi & Muduev, 2007).

The first wave was related to the deportations of the 1940s. After Chechens were sent into exile in 1944, about 60,000 of Dagestan's highland residents were forced to resettle the areas previously inhabited by the Chechens (Poljan, 2001, p. 133), including today's Chechnya and Dagestan. After Chechens were allowed to return to their homeland in 1956, Dagestanians who had moved there from the mountains were resettled once again, this time in the rural valleys of Dagestan.

A second migration took place due to the agricultural policies of Soviet authorities, who distributed huge swaths of pastureland in the valleys among cattle breeders who were originally based in the highlands. This land redistribution resulted in the relocation of agricultural workers and their families from the highlands to the rural lowlands. On many occasions, a minor migration started by small groups of agricultural workers was later augmented by many others who followed from their mountain homelands. Settlement communities of migrants, some with more than 500 households, gradually spread throughout pastoral lands. In 2010, more than 100 such settlements were reported by local authorities, some without any legal status. Their total population is estimated to be around 50,000 (Kazenin, 2012).

Finally, in the 1960s and 1970s, Soviet authorities prompted migrations in the wake of earthquakes and other natural disasters in the mountains. Then, after the collapse of the Soviet Union, both rural-to-urban and mountain-to-valley migration accelerated noticeably due to poor economic conditions in rural Dagestan, especially in the highlands. Unfortunately, official statistics only reflect data on rural-to-urban migration,

Table 2.2 Percentage of Urban Populations, 1989 and 2010

	1989	2010
Avar	30.77%	36.94%
Dargwa	31.54%	40.14%
Kumyk	47.30%	50.19%
Lak	62.25%	71.36%
Lezgian	38.02%	45.70%
Tabasaran	33.05%	44.91%

Source: Soviet population census of 1989, Russian population census of 2010.

not rural-to-rural migration from highlands to lowlands. Table 2.2 shows the proportions of the urban population among major ethnic groups according to the 1989 and 2010 censuses. Actual numbers were probably higher, as migration in the northern Caucasus was probably underrepresented by the 2010 census, since many migrants were registered in their homeland (Mkrtchyan, 2011). In the decades before the collapse of the Soviet Union, the urban population of indigenous peoples in Dagestan was also growing; for example, the Avar urban population grew threefold between 1959 and 1989 (Belozerov, 2005, p. 237).

Migrants originated from areas that were, for the most part, monolingual and often represented just one dialect of a particular language. Generally, this is not the case in the resettlement areas. However, in many instances, large groups of migrants speaking the same language or dialect have continued to live in relatively compact groups after relocating to the lowlands. A considerable number of rural settlements in the valley are inhabited by migrants from just one highland village; occasionally, people from two or three villages are intermixed in such settlements. This provides relatively favorable conditions for the preservation of their native languages.

Currently, some parts of the Dagestanian lowlands are among the rural territories of Russia with the highest population density (Èldarov, Holland, Abdulagatov, Aliev, & Ataev, 2007). This causes intensive migration out of the region into other areas of the Russian Federation. For instance, there are sizable Dargwa diasporas in the neighboring Stavropol region (49,302 in 2010) and Kalmykia (7,590 in 2010), mainly in agricultural areas, and in eastern Siberia (3,722 in Tyumen Oblast in 2010), where they are mainly employed in oil fields. Lezgians also migrated intensively to eastern Siberia (16,247 in Tyumen Oblast in 2010), while many Avars (4,719 in 2010) reside in the city of Astrakhan on the northern shore of the Caspian Sea, where they are involved in retail for the most part (for details on post-Soviet migrations from Dagestan, see Karpov & Kapustina, 2011).

2.1.2 Chechnya

Chechnya (the Chechen Republic), centrally located in the northeastern Caucasus, borders Dagestan to the east and Ingushetia to the west. Among North Caucasian republics, Chechnya is the second most populous (1,414,865 as of January 1, 2017), with Chechens comprising 95.3% of the population according to the 2010 census. Between the 1920s and 2000s, the administrative borders and ethnic composition of today's territory of the Chechen Republic underwent a number of dramatic changes, which are partly reflected in its current demographic and political situation.

During the first decade of the Soviet state, in 1922, Chechens first got their ethnic autonomous district (Rus. *avtonomnaja oblast´*). The territory of that district roughly corresponded to today's Chechen Republic without its northernmost areas. Initially, the population of the autonomous district was composed almost entirely of ethnic Chechens (94.0% according to the 1926 census). Later on, however, territories where Russian populations prevailed were incorporated into the Chechen autonomous district. In 1929, Grozny, an industrial town where Russians composed 70% of the

population and Chechens only 2% (1926 census), became a part of the autonomous district, along with some rural areas populated by Russian Cossacks. Despite the close proximity of the Chechen and Russian areas, migration between them was very uncommon at that time. Chechens resided in both the highlands and the lowlands of what is today's Chechen Republic, whereas Russians inhabited only the lowlands.

In 1934, the Chechen and Ingush autonomous districts were merged into the Chechen-Ingush District (in 1936 it became the Chechen-Ingush Autonomous Soviet Socialist Republic within the Russian Federation). This did not seriously affect the ethnic composition of the territory, because the Chechen and Ingush populations infrequently moved into each other's territories during the time they shared an administrative district, except some migration of the Ingushes, mainly officials and their family members, to the capital Grozny.

The next set of changes occurred in 1944, when the Chechens and Ingushes were accused of collaborating with the Nazis who invaded parts of the North Caucasus during World War II. Based on these accusations, Chechens and Ingushes were forcibly resettled to Central Asia. The Chechen-Ingush Autonomous Soviet Socialist Republic was dismantled. Most of its territory was absorbed into the newly established Grozny District (named after its capital, Grozny), where Russians became the majority. Some territories in the east of the former Chechen-Ingush Republic were transferred to Dagestan, and the Dagestanian population was forced to relocate there (see section 2.1.1).

In 1956, the Chechens were allowed to return to their homeland, and the following year, the Chechen-Ingush Autonomous Republic was reestablished. Most Chechens who had been living in the highlands before the exile preferred to live in the lowlands after their return, and many of those who had been living in rural areas became urbanized. This resulted in an overlap between the Russian and Chechen populations and in the depopulation of the highland areas. For instance, in the rural Shelkovskoy District to the northeast, where Chechens composed less than 1% of the population before exile, their population had grown to 19.4% by 1970. Changes in the ethnic composition of the lowlands were also due to a higher birth rate among Chechens as compared to Russians (Belozerov, 2005, pp. 231–237). The growth of the lowland Chechen population, especially in Grozny, caused a number of violent conflicts after their return from exile (Kozlov, 2006). Although the majority of Chechens returned from Central Asia as early as in 1956–1959 (Poljan, 2001, pp. 160–162), their return continued through the 1960s and 1970s, causing their population to grow even more (see Table 2.3).

After the collapse of the Soviet Union, the situation in Chechnya started to change dramatically. In 1991, Chechen separatists led by a former Soviet Army general Dzhokhar Dudaev declared the independence of the Chechen Republic, which was then given the name The Chechen Republic of Ichkeria. That caused an outburst of violence, the result of which was that the government of the Chechen-Ingush Autonomous Republic ceased to exist by the beginning of 1992, and Chechnya became predominantly controlled by separatists (and partly by unofficial militias which competed with the separatists and declared themselves pro-Russian). Ingushetia separated from Chechnya and became a republic of the Russian Federation in the same year (see section 2.1.3 for details).

Table 2.3 Major Ethnic Groups of Chechnya in 1970, 1989, and 2010

	1970	1989	2010
Chechen	499,962 (54.7%)	715,306 (66.0%)	1,205,551 (95.1%)
Russian	329,701 (36.1%)	269,130 (24.8%)	24,382 (1.9%)
Ingush	14,543 (1.6%)	25,136 (2.3%)	1,296 (0.1%)
Kumyk	6,865 (0.8%)	9,591 (0.9%)	12,221 (1.0%)
Noghay	5,503 (0.6%)	6,885 (0.6%)	3,444 (0.3%)
Avar	4,196 (0.5%)	6,035 (0.6%)	4,864 (0.4%)

Note: Unlike other regions, we provide census results from 1970 for Chechnya because the ethnic composition reflected in the 1959 census was questionable, due to the ongoing return of Chechens and Ingushes from Central Asia in exactly that period.
Source: Soviet population censuses of 1970 and 1989, Russian population census of 2010.

The military conflict between separatists and the Russian Army from 1994 to 1996 (the First Chechen War) did not bring Chechnya under the control of the Federal government. The Kremlin was more successful in the Second Chechen War (1999–2002), which resulted in the defeat of the separatists and the authorization of a pro-Russian government, wherein former separatist commanders played, and still play, the main role. After the Second Chechen War, the Kremlin started large-scale investments in restoring the physical and social infrastructure of the Chechen Republic, which was declared a separate region of the Russian Federation in its Constitution of 2003.

The military conflicts had several consequences for the population and the ethnic composition of the Chechen Republic. Foremost, they resulted in large-scale losses of life and many refugees. During the First Chechen War, human rights activists were the main source of data for civilian casualties. The Russian human rights center *Memorial* reports that 30,000 to 40,000 civilians were killed in Chechnya from 1994 to 1996. The Federal government reported 17,000 losses among separatists. Assessments of civilian losses during the second conflict vary between 1,000 (Russian officials) and 25,000 (Amnesty International). Russian assessments of the separatists' losses from that war vary between 10,000 and 15,000. The number of refugees from both wars amounted to tens of thousands, although exact estimates made by military observers vary. The highest concentration of refugees was in the neighboring republic of Ingushetia. Until 2010, most refugees who returned to Chechnya came from Ingushetia (see section 2.1.3). The number of refugees who went to other regions is hard to assess due to the unreliability of migration statistics.

The years of separatism and war turned Chechnya into an ethnically homogeneous region for the first time in its history. This was reflected in the 2002 and 2010 census results (for 2010, see Table 2.3). The decrease in the Russian population in 2010 was especially noticeable when compared to 1989. Unlike the Chechens, most Russians who left the area during wartime never returned. The number and proportion of Ingushes also decreased after 1989. At the time of the Chechen-Ingush Autonomous Republic, Ingushes

in Chechnya were found only in the present-day city of Grozny. Their numbers in today's Chechnya have dwindled from 25,136 in 1989 to 1,296 in 2010, which shows the intensity of their emigration, mainly to Ingushetia. Since the split between Chechnya and Ingushetia in 1992, no border between the two republics has been officially agreed upon. However, the unofficial border was configured so that no predominantly Ingush areas are included in Chechnya. Nevertheless, in 2018 debates concerning the borders of the republics intensified. After the head of Chechnya, Ramzan Kadyrov, and the head of Ingushetia, Junus-Bek Evkurov, signed a border demarcation agreement in October 2018, more than 10,000 people rallied in Ingushetia, protesting against the secretive nature of the agreement and the particulars of how the border was settled.

There are only three minorities in Chechnya with more than 2,000 people: Kumyks, Noghays, and Avars. Kumyks (12,221 in 2010) reside mainly in some villages in eastern Chechnya. Noghays (3,444 in 2010) reside in the northern part of Chechnya, close to the Dagestanian Noghay populations. The Noghay population reduced twofold from 1989 to 2010 because, during wartime, they fled to other regions of Russia, including Dagestan, Astrakhan, and Tyumen, where large Noghay communities already existed. Finally, Avars, though mostly found in Dagestan, live in some eastern villages in Chechnya (4,864 in 2010). They are mainly descendants of the Avars relocated to Chechnya by Soviet authorities after the Chechens were exiled in 1944 (see section 2.1.1). Only a very small part of the Avar population stayed in Chechnya after the Chechens returned in 1957. In Borozdinovskaya, the village with the highest concentration of Avars today, Avars experienced severe pressure from some paramilitary groups in 2005, when 11 local residents, still missing today, were taken away by force from the village. After that, nearly 1,500 Avars migrated from Borozdinovskaya to Dagestan.

All in all, in less than a century, the territory of today's Chechen Republic has transitioned from a multiethnic region with separate residential areas for each major ethnic group to a region with interethnic mixing, and, most recently, to an almost entirely monoethnic region.

The post-war restoration of Chechnya improved the living conditions, both in towns and in major rural settlements, to the effect that post-war urbanization was not too prevalent (possibly also because it was partially controlled by local authorities). Although the population of the regional capital, Grozny, which was totally rebuilt after the wars, grew 38% between 2002 and 2017, Chechnya still has one of the highest proportions of rural residents in Russia (65% in 2016). High population growth is another demographic characteristic of Chechnya. Its population grew 11.5% between the 2010 census and January 1, 2017. This may be due in part to the high birth rate which was 23‰ in 2016, almost double of that of Russia as a whole.

2.1.3 Ingushetia

Ingushetia (the Republic of Ingushetia) has the smallest territory of all Russian regions in the northern Caucasus (3,628 square km). As a distinct region of Russia, its present borders were established in 1992. Before that, between the 1920s and 1950s, the territory

of today's Ingushetia had changed several times. From 1957 to 1991, it was a part of the Chechen-Ingush Autonomous Soviet Socialist Republic within the USSR.

When Chechnya claimed independence in the 1990s, Ingushetia remained a part of the Russian Federation. Almost immediately after its formation, the Republic of Ingushetia received two dramatic influxes of people. Both were caused by nearby incidents of mass violence. One began after a conflict in North Ossetia. By the time of the Soviet Union's collapse, tens of thousands of Ingushes lived in North Ossetia (32,723, according to the census of 1989, but that is purported to be an underestimate). More than 50% of Ingushes in North Ossetia were concentrated in the Prigorodny District, a territory that belonged to the Chechen-Ingush Autonomous Republic before 1944. Upon their return from exile in 1957, Ingushes were allowed to reside in their homeland, part of which remained within the territory of North Ossetia. In October 1992, a bloody conflict between the Ossetic and Ingush populations broke out in Prigorodny District, with the following casualties: 583 dead, 939 injured, and 261 missing (based on the Russian authorities' reports). As a result of that conflict, almost the entire Ingush population of North Ossetia was forced to move to Ingushetia. Estimates for the number of refugees ranged from 30,000 to 60,000, depending on the source. Although a fraction of those refugees has since returned to North Ossetia (see section 2.2.1), an uncertain number of refugees and their adult children still remain in Ingushetia.

In addition to the conflict in North Ossetia, the other major cause of people relocating to Ingushetia was extensive violence in Chechnya, whose declaration of independence resulted in two large wars, described in section 2.1.2. Almost all the Ingushes who lived in the capital, Grozny (21,346 according to the 1989 census), fled to Ingushetia when Grozny became the center of the rebellion in Chechnya. Many Chechen refugees fled to Ingushetia as well. According to the Danish Refugee Council, the total number of refugees from Chechnya to Ingushetia was nearly 106,000 in 2002. In the same year, the Russian census registered 95,403 Chechens in Ingushetia. Afterward, Chechen refugees either returned to Chechnya or relocated to other parts of Russia. From 2002 to 2010, the number of Chechens in Ingushetia shrank fivefold. However, very few, if any, Ingushes who fled Chechnya in the 1990s have returned to Ingushetia (see Table 2.4 for changes of ethnic proportions in Ingusgetia between 1959 and 2010).

Table 2.4 Major Ethnic Groups of Ingushetia in 1959, 1989, and 2010

	1959	1989	2010
Ingush	44,634 (40.6%)	138,626 (74.5%)	385,537 (93.5%)
Chechen	5,643 (5.1%)	19,195 (10.3%)	18,765 (4.5%)
Russian	51,549 (46.9%)	24,641 (13.2%)	3,215 (0.8%)

Note: For 1959 and 1989, the numbers are given for those administrative units of the Chechen-Ingush Autonomous Republic whose territory was mainly included in the Republic of Ingushetia in 1992.
Source: Soviet population censuses of 1959 and 1989, Russian population census of 2010.

Overall, immigration has considerably increased the total number of Ingushes who now permanently reside in Ingushetia, although it is hard to assess how many of them are actually migrants. Migration of Russians out of the region and a large disparity between their birth rates and those of local Ingushes, who have the highest birth rate in Russia, have made Ingushetia nearly monoethnic; Ingushes compose 93.5% of the population according to the 2010 census. The overall population growth rate is very high.

At present, the lowest population density in Ingushetia is found in the southern highlands, where less than 5% of the total population resides. The highest population density is found in the central part of the region, where towns and rural settlements form an agglomeration along the federal highway. The current proportion of the urban population is estimated to be 41.8% (2017). However, the actual level of urbanization is somewhat lower, as many districts of the major towns (Nazran, Malgobek, Karabulak) consist mostly of private one-family buildings and generally have rural infrastructure. Only the newly built capital of Ingushetia, Magas, has full-fledged urban infrastructure. Its population, however, is estimated at only 7,818 in 2017.

2.2   Northwestern Caucasus

The northwestern Caucasus includes four republics within the Russian Federation: North Ossetia, Kabardino–Balkaria, Karachay–Cherkessia, and Adygea. Compared to the northeastern Caucasus, these republics have populations made up of a higher proportion of Russians and a lower proportion of practicing Muslims. All in all, the ethnic composition and religion of these republics contrasts less with the rest of Russia than of those in the northeast.

Sometimes, Stavropol Krai and Krasnodar Krai are also considered part of the northwestern Caucasus. We will not discuss them in this chapter, as Russians constitute the overwhelming majority of their population.

2.2.1 North Ossetia

North Ossetia (the Republic of North Ossetia–Alania), centered in the North Caucasus, is the only region of the North Caucasus which has a direct connection to Georgia and South Ossetia by road. It shares borders with Ingushetia to the east, Kabardino–Balkaria to the west, Stavropol to the north, and Georgia and South Ossetia to the south. The total population amounted to 703,262 in 2017. The urban population in 2017 constituted 64.2% of the total population, considerably higher than in other republics of the North Caucasus. Most of the urban population is concentrated in the capital, Vladikavkaz (324,836 in 2017).

The religious composition of North Ossetia is also unique among North Caucasian republics, as the proportion of Muslims is much lower than 50%. The two major ethnic

groups, Ossetians (64.5%, according to the 2010 census) and Russians (20.6%, according to the same census), are predominantly Christian. There are some Ossetian minorities who historically identified as Muslim, but their numbers are small.

Russians and Ossetians have a long history of living together, since Vladikavkaz, as early as the nineteenth century, was the North Caucasian administrative center of Tsarist Russia. According to the 2010 census, more than 54% of the Russian population of the republic was concentrated in Vladikavkaz; in rural districts, Ossetians make up almost 90% of the population. The northern part of the republic (the town of Mozdok and the surrounding rural areas) is the only area where Russians are close to a majority (49.9% in 2010). That area was added to North Ossetia in 1944 and has a very different ethnic composition from the rest of the republic (see below in this section about the Kumyks living there). Changes in the ethnic composition of North Ossetia between 1959 and 2010 are shown in Table 2.5.

A relatively small decrease in the Russian population from 1970 to 2010 is explained by moderate rates of emigration from North Ossetia, much lower than in the republics of the northeast. The increased proportion of Ossetians could be due to less emigration among them than among the Russians and, partly, to immigration (differences in birth rate are probably not relevant, as they are negligible between the two groups).

In the late 1980s and early 1990s, Ossetians migrated to their homeland primarily from Georgia. According to the 1989 census, 164,055 Ossetians lived in Georgia, at that time including the South Ossetic Autonomous Oblast (SOAO), which was part of the Georgian Soviet Socialist Republic. After 1989, when a military conflict between Ossetians and Georgians broke out, Ossetians almost completely left the regions of Georgia that were outside the SOAO. They migrated to both North Ossetia and South Ossetia. Further migration to North Ossetia took place from South Ossetia, where the living conditions were very harsh in the 1990s. Since not all of those relocations were registered, it is impossible to calculate the exact increase in the Ossetian population in North Ossetia they have led to. Russian officials estimated the number of Ossetian refugees in North Ossetia at about 30,000 by 1993 (*Rossijskaja gazeta*, March 10, 1993).

Table 2.5 Major Ethnic Groups of North Ossetia in 1970, 1989, and 2010

	1970	1989	2010
Ossetian	215,463 (47.8%)	334,876 (53.0%)	459,688 (64.5%)
Russian	178,654 (39.6%)	189,159 (29.9%)	147,090 (20.6%)
Ingush	6,071 (1.3%)	32,783 (5.2%)	28,336 (4.0%)
Kumyk	3,921 (0.9%)	9,478 (1.5%)	16,092 (2.3%)
Armenian	12,012 (2.7%)	13,619 (2.2%)	16,235 (2.3%)
Georgian	8,160 (1.8%)	12,284 (1.9%)	9,095 (1.3%)

Source: Soviet population censuses of 1970 and 1989, Russian population census of 2010.

Their subsequent number is even more difficult to determine because of "pendulum migration" between North and South Ossetia which often stays unregistered.

The two most populous minorities of North Ossetia are the Kumyks and Ingushes. Ingushes are concentrated in the Prigorodny District between Vladikavkaz, its suburbs, and the border with Ingushetia. Their numbers fluctuated seriously in the post-Soviet era because of the conflict in the Prigorodny District in 1992 (see section 2.1.3). However, the 2010 census showed that the number of Ingushes, who were almost completely gone in 1992, had almost reached their pre-conflict level from 18 years earlier. That is partially due to the return of Ingush refugees and partially to the much higher birth rate among the Ingush population than among the Ossetian and Russian populations (Belozerov, 2005, pp. 131–137). Another Muslim minority, the Kumyks, are concentrated in the northern reaches of North Ossetia. In the Mozdok District, they composed 18.6% of the population (according to the 2010 census), outnumbered only by Russians.

Georgians and Armenians are concentrated in Vladikavkaz (Armenians are also found in the northern areas, and Georgians in some far southern villages, close to the highway connecting North Ossetia with Georgia).

2.2.2 Kabardino–Balkaria

Kabardino–Balkaria is a republic in the center of the North Caucasus with an estimated population of 864,454 (2017). Approximately 40% of its territory is covered by mountains, the rest being foothills and lowlands. Three major peoples are found there: Kabardians, Russians, and Balkars.

Kabardians, whose proportion of the population in the region is well above 50%, belong to the eastern branch of the Circassian ethnic group. Kabardians inhabit the rural valleys in the west and southeast of the region almost exclusively. They also compose more than 90% of the population in the western town of Baksan. Kabardians started migrating to towns outside their traditional areas in the mid-twentieth century. In the 2010 census, less than 1% of Kabardians identified themselves with the broader ethnonym, "Circassian" (see also chapter 9).

Balkars, a Turkic people, had mainly inhabited highlands and neighboring foothills before 1944. In that year, they were forced to move to Central Asia by Soviet authorities, who accused them of collaboration with the Nazis during World War II. After returning to their homeland in 1957, the Balkars have maintained a relatively stable geographic range of residence. However, some Balkars, who had been living in the mountains before exile, resettled in Balkar villages around the capital, Nalchik. Later on, a small number of Balkars moved to ethnically mixed lowlands in northeastern Kabardino–Balkaria.

Russians mainly reside in two areas: Nalchik and the northeastern regions, including the town of Prokhladny. In recent decades, the proportion of Russians in Kabardino–Balkaria has not decreased as sharply as in the northeastern Caucasus; this is due to less drastic rates of emigration and a narrower gap in the birth rate between the Russians and the indigenous populations.

Table 2.6 Major Ethnic Groups of Kabardino–Balkaria in 1959, 1989, and 2010

	1959	1989	2010
Balkar	34,088 (8.1%)	70,793 (9.4%)	108,577 (12.6%)
Kabardian	190,284 (45.3%)	363,494 (48.2%)	490,453 (57.0%)
Ossetian	6,442 (1.5%)	9,996 (1.3%)	9,129 (1.1%)
Russian	162,586 (38.7%)	240,750 (31.9%)	193,155 (22.5%)
Ukranian	8,400 (2%)	12,826 (1.7%)	4,800 (0.6%)

Source: Soviet population censuses of 1959 and 1989, Russian population census of 2010.

Other ethnic groups comprise less than 10% of Kabardino–Balkaria's total population. Meskhi Turks are among the minorities whose number has been growing in recent decades (4,162 in 1989, 11,053 in 2002, and 13,965 in 2010). Expelled by Stalin from Georgia in 1944, they were forced to move to Central Asia, and some returned to Kabardino–Balkaria after their exile was lifted in 1957 (Poljan, 2001, pp. 175–179). Another group migrated from Central Asia around the collapse of the Soviet Union. Today, they mainly reside in rural parts of eastern and northeastern Kabardino–Balkaria.

Since the 1960s, internal migration in Kabardino–Balkaria has gone in two directions. First, Kabardians and Balkars tend to migrate to Nalchik and the neighboring rural settlements. The proportion of Kabardians in Nalchik changed from 18.8% in 1970 to 49.25% in 2010, and the proportion of Balkars, from 2.7% to 12.16%. Some rural settlements around Nalchik are now multiethnic, but others remain monoethnic (with either Balkars or Kabardians). Second, Kabardians and Meskhi Turks tend to migrate to the northeastern parts of the republic, where Russians are gradually leaving rural areas. The central town of the northeast of the republic, Prokhladny (57,879 in 2017), formerly a Russian Cossack settlement, has become remarkably diverse in recent decades in terms of its ethnic composition (for details on changes of "ethnic borders" due to migration within the region, see Babich & Stepanov, 2009). All in all, however, post-Soviet rural-to-urban migration was less intense in Kabardino–Balkaria compared to the northeastern Caucasus, possibly due to the higher current level of agricultural development creating more job opportunities in the rural part of this republic (see Table 2.6 for changes in the ethnic composition of the region between 1959 and 2010).

2.2.3 Karachay–Cherkessia

Karachay–Cherkessia (Karachai–Cherkessia) is a republic in the northwestern part of the Caucasus. Until 1990, when it became a separate member of the Russian Federation, it was part of Stavropol Krai, where Russians constitute a majority. The total population of the republic amounted to 466,432 in 2017. Almost 80% of its territory is covered by

mountains and foothills. The major peoples are the Karachays, a Turkic people linguistically close to the Balkars of Kabardino–Balkaria; the Circassians, represented here mainly by an ethnic subgroup closely related to Kabardians; the Abazins (Abaza), who are related to the Circassians; and the Noghays, a Turkic people also present in Dagestan, Stavropol, and some other regions of Russia. None of these four major peoples constitute a majority of the total population of the region.

Most of Karachay–Cherkessia's current territory was first integrated by Soviet authorities in 1922 into a single administrative unit, which dissolved into several parts in 1926, and reunited in 1957 as an autonomous district within the Stavropol region. After the collapse of the Soviet Union, many proposals to split the region into two or three republics were put forward, but none of them were implemented, despite considerable interethnic tension in the region at the time (Kazenin, 2009b, pp. 121–128). With the exception of the regional capital, Cherkessk (originally founded as a Cossack village), and a small number of rural settlements with high ethnic diversity, the people of Karachay–Cherkessia still live in separate communities, where most settlements are monoethnic and a few are inhabited by just two ethnic groups.

Karachays were concentrated in highlands to the south and foothills to the east before their deportation to Central Asia in 1943. After their return in 1957, they only partially went back to their former homeland. Many of them chose to settle in rural areas previously inhabited almost exclusively by Russians, thereby expanding into rural territories in the northeastern and western parts of the region (Belozerov, 2005, pp. 108–130). Karachays also constitute a majority in the town of Karachayevsk, which was the center of the separate Karachay autonomous district during its short existence between 1926 and 1943. Their proportion in Cherkessk has been increasing (from 6.2% in 1979 to 16.2% in 2010).

Circassians inhabit the northwestern part of the republic. Rural Circassians reside primarily in ethnically homogeneous villages. Their proportion in Cherkessk grew from 6.4% in 1979 to 13.0% in 2010.

Russians still constitute a majority in Cherkessk. Their migration from rural settlements shared with Karachays continues today. For instance, in the northeastern Prikubansky District, the proportion of Russians and Karachays was 28.4% and 47.0%, respectively, in 1979, 18.5% and 56.2% in 2002, and 17.2% and 75.69% in 2010. This trend is part of a general tendency for Russians to move from rural to urban areas in the south of the country (Nefedova, 2015). Aside from urbanization, emigration out of the republic is the main reason for the ongoing decrease in the Russian population (see Table 2.7). Differences in the birth rates of Russians and other peoples in Karachay–Cherkessia are not currently significant.

Abazins are concentrated in several enclaves scattered throughout the region and in Cherkessk, whereas Noghays inhabit the northwest, close to Circassians. In rural areas, Abazin and Noghay villages tend to be monoethnic. In 2007, Abazins and Noghays were made titular ethnic groups in two respective districts. Nearly half of Abazin villages were included in the Abazinsky (Abaza) District, and the Noghaysky (Noghay) District constitutes almost the whole area of Noghay settlement in the region (Kazenin, 2012, pp. 104–127).

Table 2.7 Major Ethnic Groups of Karachay–Cherkessia in 1959, 1989, and 2010

	1959	1989	2010
Abazin	18,159 (6.5%)	27,475 (6.6%)	36,919 (7.7%)
Circassian	24,145 (8.7%)	40,241 (9.7%)	56,466 (11.8%)
Karachay	67,830 (24.4%)	129,449 (31.2%)	194,324 (40.7%)
Noghay	8,903 (3.2%)	12,993 (3.1%)	15,654 (3.3%)
Russian	141,843 (51.0%)	175,931 (42.4%)	150,025 (31.4%)

Source: Soviet population censuses of 1959 and 1989, Russian population census of 2010.

Like Kabardino–Balkaria, Karachay–Cherkessia experienced less rural-to-urban and highland-to-lowland migration than the northeastern republics did after the collapse of the Soviet Union. However, the migration of Russians out of rural areas, as well as the migration of all ethnic groups to the regional capital still changed the ethnic composition of this area dramatically in the post-Soviet period.

2.2.4 Adygea

Adygea (Adyghea) is the westernmost republic of the North Caucasus; it borders Krasnodar Krai, a major region of Southern Russia. Adygea was established in 1937 as an autonomous district within Krasnodar Krai, with borders similar to the current ones. In 1990, Adygea separated from Krasnodar Krai and became a separate region of the Russian Federation.

Unlike other republics of the North Caucasus, the titular ethnic group of Adygea, the Circassians (Adyghe), are actually a minority in its population. They have never amounted to even a quarter of Adygea's population, whereas the proportion of Russians has stayed above 60% (see Kabuzan, 1996, p. 117; Kazenin, 2009b, pp. 14–20). In fact, at the time of its formation, Adygea combined predominantly Circassian and Russian territories, and the population of the latter outnumbered the population of the former. Differences between parts of the region in ethnic composition have been preserved throughout the republic up to the present day.

When considering the ethnic composition of Adygea, some shifts in ethnic terminology should be taken into account. In the 2010 census, the Circassian population was given the option to register either under the name "Circassian" (Rus. *čerkesy*) or "Adyghe" (Rus. *adygejcy*); only the latter was officially acknowledged during most of the Soviet era. The ethnonym "Adyghe" was initially applied by Soviet authorities to all Circassians outside Kabardino–Balkaria (Circassians there were classified as Kabardians). Starting in the 1930s, the use of "Adyghe" was restricted to the Circassians of Adygea. Before the 2010 census, some local Circassian activists advocated for the return of the term "Circassian" in place of "Adyghe" as a symbolic reunification of the entire Circassian

Table 2.8 Major Ethnic Groups of Adygea in 1959, 1989, and 2010

	1959	1989	2010
Russian	235,539 (72.8%)	293,640 (68.0%)	270,714 (61.5%)
Adyghe	65,955 (20.4%)	95,439 (22.1%)	107,048 (24.3%)
Ukrainian	9,461 (2.9%)	13,755 (3.2%)	5,856 (1.3%)
Circassian	–	–	2,651 (0.6%)
Armenian	4,659 (1.4%)	10,460 (2.4%)	15,561 (3.5%)
Kurd	–	–	4,582 (1.0%)

Source: Soviet population censuses of 1959 and 1989, Russian population census of 2010.

population of the Caucasus. However, rather few (less than 3,000) Circassians actually used the opportunity to change their official self-identification in the census, so their choice did not affect the apparent ethnic composition of the republic (see Table 2.8 for changes in the ethnic composition of Adygea between 1959 and 2010).

Circassians/Adyghes are a majority in the west of the republic, close to the city of Krasnodar. There they live in the town of Adygeysk, considered a satellite of Krasnodar, and in a number of rural settlements. Some Circassians were moved to their current location in the early 1970s, before their former territory was flooded by the Krasnodar Water Reservoir built in 1973. Circassians also live in the northeastern part of Adygea, along the west bank of the Kuban River. The central part of Adygea, which separates the two Circassian enclaves, is predominantly inhabited by Russians. That area includes Adygea's capital, Maykop, and two rural districts where the number of Circassians is especially low (less than 5% in 2010).

Comparison of census data from 1959, 1989, and 2010, for Adygea shows that ethnic proportions have changed little since 1959. The lack of noticeable changes may be due to the relatively low emigration rates among other North Caucasian groups and to insignificant differences in the birth rates between the Russians and the Circassians. At the end of the Caucasian Wars in the nineteenth century, many Circassians had migrated to Turkey and the Middle East; possible repatriation of their descendants was widely discussed after the collapse of the Soviet Union, but the actual immigration of Circassians to their homeland in Adygea has been extremely modest and has not affected the total proportion of Circassians in the region.

The low proportion of Circassians in Adygea has caused political tensions in the post-Soviet era, as Russian political activists have protested against the high proportion of Circassians among local authorities and against some laws of the republic, which they claimed discriminate against the Russian population (including the requirement for presidential candidates to speak Circassian and the system of parliamentary representation which, in practice, guaranteed Circassians a majority in the upper house of the local parliament). In the early 2000s, most of these disputed regulations, including the language requirement, were abolished (Kazenin, 2009b, p. 31).

Migration within Adygea has not been intensive in recent decades. Only Maykop has attracted a considerable population influx. Its total population grew from 82,135 in 1959 to 144,249 in 2010, and the proportion of Circassians in Maykop changed from 3.4% in 1959 to 18.9% in 2010. The rather low proportion of Circassians in the capital of the republic, as well as its more modest population growth compared to the capitals of other republics in the North Caucasus, can be explained by its proximity to Krasnodar, where the economy and urban infrastructure are much more developed than in Maykop. Therefore, Krasnodar is a more attractive urban destination for the rural citizens of Adygea than its own capital is. Yet not all Circassians of Krasnodar Krai have migrated from Adygea. Of the 13,800 Circassians registered in the 2010 census of the Krasnodar District, most lived in its areas already inhabited by Circassians for several centuries. There are, however, reasons which suggest that the present-day migration of both Russians and Circassians from Adygea to the Krasnodar District is understated. Another reason for relatively low migration into the capital of Adygea may have to do with the advantages that the climate of the lowlands offers to agricultural work. Highly developed agriculture in both Circassian and Russian areas of the republic forestalls urbanization, at least to some degree. At present, the urban population constitutes 47.31% (2017), which is much lower than in Russia as a whole.

Apart from Maykop, there are only two areas in Adygea where considerable ethnic intermixing is found—first, along the westernmost margin of the republic, immediately adjacent to Krasnodar. Although Circassians constitute the majority there, the Russian population is motivated to relocate there, as home prices are lower there than in Krasnodar, and the distance from the city is short. Second, Russians and Circassians share some rural parts of northeastern Adygea, where they have been neighbors for more than a century.

Among the ethnic minorities present in Adygea, Armenians are the largest. They can be considered a subgroup of the much more numerous Armenian population of Southern Russia, mainly living in Krasnodar Krai. Kurds migrated to Adygea after the collapse of the Soviet Union, mainly from Central Asia, where they had been deported during Stalin's reign (Poljan, 2001, p. 137). They inhabit a small number of rural settlements in Adygea. Ukrainians, who initially moved to agricultural areas of Adygea in the nineteenth and early twentieth centuries, are almost entirely assimilated with Russians today.

Further Reading

Fundamental studies of ethnic composition and other aspects of demography of North Caucasus are almost all in Russian. This section offers a brief survey of the most informative among them.

Belozerov (2005) gives an overview of interethnic differences in birth rates, paths of migration, and proportions of urban population, mainly focusing on the western part of the North Caucasus. Changes in the ethnic composition in various parts of the

region in the last decades of the USSR and the early post-Soviet years are also considered.

Kabuzan (1996) offers a historical account of changes in the ethnic composition of the North Caucasus in the Russian Empire and in the Soviet era, making use of a large set of documents.

Karpov and Kapustina (2011) study migration of rural (mainly highland) populations of the North Caucasus to the lowlands within and away from their native regions. Social transformations accompanying those migration processes are also discussed.

Karpov (2017) provides a study of the formation of administrative units (republics, etc.) in the North Caucasus in the 1920s and 1930s. The book accounts for the emergence of a system of administrative borders which, for the most part, has been preserved in the North Caucasus still today and which most conflicts mentioned in this chapter are related to.

PART II

NAKH-
DAGESTANIAN
LANGUAGES

CHAPTER 3

··

NAKH-DAGESTANIAN LANGUAGES

··

DMITRY GANENKOV AND TIMUR MAISAK

THIS chapter provides an overview of the basic properties of the languages of the Nakh-Dagestanian (East Caucasian) family. Given the size of the family, we cannot cover even the most typical features in full here, let alone describe details of the variation that exists. Likewise, we cannot do full justice to all individual languages or even branches within the family and must instead confine the discussion to occasional mentions of languages and branches here and there. The goal of this chapter is to complement the body of previously published surveys of the family and its branches, such as van den Berg (2005b), Bokarev and Lomtatidze (1967), Klimov and Alekseev (1980), Smeets (1994), Alekseev (1998b), Hewitt (2004), and Job (2004), and to provide a state-of-the-art update on the major issues in the grammar of Nakh-Dagestanian. Where appropriate, we refer the reader to other chapters in this volume or to existing family- or branch-wide overview studies of specific phenomena. For reasons of space, however, we do not provide references to individual grammatical descriptions, except when citing examples from the literature. Examples without references are drawn from our own fieldwork.

3.1 INTRODUCTION

In this section, we discuss the diachronic relationships and geographical distribution of languages within the Nakh-Dagestanian family (3.1.1). We then proceed to an overview of available historical sources (3.1.2), current sociolinguistic situation (3.1.3), and the history of research (3.1.4) on languages of the family.

3.1.1 Structure of the family

The Nakh-Dagestanian languages are spoken in the eastern Caucasus (and for this reason they are also commonly labeled "East Caucasian"). The majority of these languages are located in the Republic of Dagestan, Russian Federation. The Republics of Chechnya and Ingushetia (both also belonging to the Russian Federation) are home to Chechen and Ingush, respectively. Northern regions of Azerbaijan and eastern parts of Georgia bordering Dagestan and Chechnya also host Nakh-Dagestanian-speaking communities. Smaller communities live in Turkey, Kazakhstan, and Kyrgyzstan and are dispersed more widely across the Russian Federation.

The family is divided into four accepted branches: Nakh, Avar-Andic-Tsezic, Lezgic, Dargwa, and two family-level isolates—Lak and Khinalug—each constituting a separate branch (see Figure 3.1).

Lack of clear dialectal divisions and dearth of historical reconstructions of Nakh-Dagestanian are the main reasons why the internal composition of the family, especially on the Dagestanian side, is still subject to debate. The groupings presented here are similar, but not identical, to the groupings discussed in chapter 1, and this is a reflection of an ongoing debate in Dagestanian language studies. In particular, it is not clear whether, in fact, the Nakh branch is opposed to a Dagestanian branch including all other groups, as suggested by the family's name, or whether the root node splits off into several branches as illustrated in Figure 3.1. It may be that the Tsezic languages are a separate branch, while Lak has been grouped together in a branch with Dargwa. Khinalug, spoken in one village in Azerbaijan, was formerly included in the Lezgic branch, but the affinity between the two is now considered to be areal rather than genetic in nature.

The Nakh languages are spoken in Chechnya (Chechen), Ingushetia (Ingush), and Georgia (Batsbi, also known as Tsova-Tush). The nine Avar-Andic languages—Avar, Akhvakh, Andi, Bagvalal, Botlikh, Chamalal, Godoberi, Karata, and Tindi—are spoken
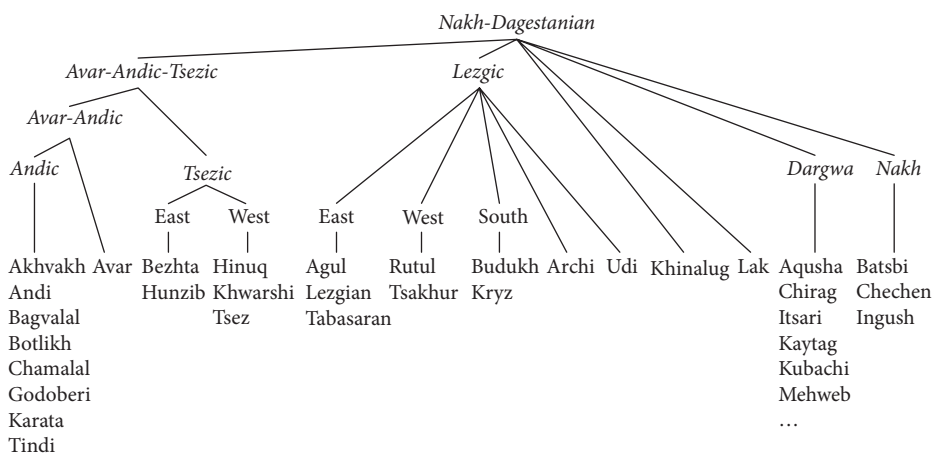

FIGURE 3.1. Nakh-Dagestanian languages.