Introduction
Maria Polinsky

1.   Navigating the Area

The Caucasus is a relatively small landmass 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 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 the most well-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 2011, 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 of 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.[1,2] Two of these 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.[3] The south is where languages of the Kartvelian (South Caucasian)

---

[1] See Comrie (2005) for the terminological distinction between "languages of the Caucasus" and "Caucasian languages," and see also Chapter 1.

[2] 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; 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.

[3] 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.

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 (see Chapters 3 and 8), the languages traditionally called Dagestanian do not—this term reflects common geography rather than early branching in the history of the family (see Chapters 1 and 3, for more discussion).

Researchers looking for long-range linguistic comparisons place Kartvelian languages in the Nostratic family (Illich-Svitych 1971; Bomhard 2008, a.o.) 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 (see also Chapter 1). 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: 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 (Chapter 13). 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.

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 amongst languages of the area has always been very productive. At some point, researchers were even tempted to propose the concept of the Caucasian *Sprachbund* (Klimov 1978; Klimov & Alekseev 1980; Chirikba 2008; 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.[4] 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 1998, 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 & Èfendiev (2001), and Zabitov (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 Potto 1887-89; Baddeley 1908, 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 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

---

[4] 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.

Georgian Bagrationi dynasty accepted the inclusion of their lands in the Russian empire as the lesser of two evils.[5]

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: 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 (Daniel et al. 2010; Belikov 2011; 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.

Second, despite the fact that many censuses indicate large numbers of speakers for certain languages (see Chapter 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).[6] The growing dominance of Russian underscores the urgency of studying the languages of the northern Caucasus; the often-times 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

---

[5] 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).

[6] 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.

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 (Slezkine 1994; Blauvelt 2013). 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, etc. 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.[7]

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 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 sub-groups (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.

---

[7] 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.)

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 systems, which are more complex than the usual masculine-feminine distinction.[8] 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.

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 2008: 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

---

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

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',[9] Mehweb Dargwa *k'amp'it'* 'candy', etc. 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 (but see Chapter 13, on prepositions in Indo-European languages of the area). 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 41specifically 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: 75). Thus, Abkhaz is relatively unusual in that regard.[10]

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) below. 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 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 Manetta 2011; Kural 1997, a.o.). A hallmark of head-final languages, complex predicates, formed from a lexical component and a light verb such as 'be' (for intransitives) and 'do' (for transitives), 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

pe?   ud-u-na,   mert: aq'-u-na   iži-di,   fajš-u-na,
chicken.ABS   tear-PFV-CONV 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   pe?-ela-k-as.
become-IPFV-CVB-COP   DEM.PROX chicken-OBL-SUB.CONT-ELAT

---

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

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

'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;[11] 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 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, both in terms of 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.

---

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

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).[12] 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 1987b; Klimov 1973, and see footnote 12). 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 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

---

[12] Using more idiosyncratic criteria, Klimov & 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.

linked to subject antecedents and overt pronouns to antecedents in lower structural positions (Carminati 2002, 2005). Thus, in the Spanish example below, 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 et al. 2016: 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 object 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   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 (Kuroda 1965; Nishigauchi 1990; Shimoyama 2006, Haspelmath 1997). 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, etc. Usually the bare forms have the interrogative interpretation. Consider the following paradigm from Svan (David Erschler, pers. comm.):

	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

Table 1 Indeterminate expressions in Svan

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 2001: 165-167 for Bagvalal; Polinsky 2015b for Tsez) and in Armenian and Ossetic (Haspelmath 1997: 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 2009; Wier 2011).[13]

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 below, 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

---

[13] 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 & Lander 2017; Testelets 2009).

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 and 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. Furthermore, as with all overviews, certain things have been omitted. For more on the features shared across languages of the Caucasus, see Chirikba (2008), Klimov (1978), and further references therein.[14]

    A morphological optative—the modal form that expresses wishes, desires, potentialities, or hopes—is found in almost all of the area's languages. The example below highlights Ancient Greek to illustrate another common property of morphological optatives: co-occurrence with a particular 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.'

---

[14] See also Chapters 1 and 3 for a discussion of properties shared across Nakh-Dagestanian languages.

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 2008; Dobrushina 2011; Dobrushina et al. 2013).[15] Consider examples from the three indigenous families, as well as some other languages of the area (and see Dobrushina 2011 for more examples from the Nakh-Dagestanian family):

(6) a. Adyghe

    qə-š'-ere-č̣ʼəx   qeʁaǯe-xe-r.

    DIR-LOC-OPT-grow   flower-PL-ABS

    'Let flowers grow here!' (Kuznetsova 2009: 291)

  b. Georgian

    man   unda   gadac'eros   es   c'eril-i.

    3SG.ERG MOD   3SG.copy.OPT   DEM   letter-NOM

    'He needs to copy this letter.' (Cherchi 1997: 260)

  c. Lezgian

    wa-z   allah-di   hamišan ŭʂret   gu-raj.

    2SG-DAT Allah-ERG   always   help   give-OPT

    'May God always help you.' (Haspelmath 1993: 151)

  d. Kumyk

    tez   jaz   bol-ʁaj   e-di.

    soon summer be-OPT   AUX-PST

    'I wish summer would come soon.' (Dobrushina 2011: 104)

  e. Judeo-Tat

    soχ-o-m.

    do-OPT-1SG.PST

    'Let me do it!' (Chapter 13: PAGE)

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: 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.'

---

[15]Chirikba (2008: 52) refers to this category as the 'potential.'

They presuppose an object that is affected by the action, and the medium (instrument) of the respective action. In more familiar languages, the person 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: 58-59).[16] 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)   caλi-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.

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

---

[16]Klimov (1978: 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) below, the subject is in the absolutive, and the agreement on the verb reflects an intransitive pattern; the instrument is expressed by a PP (*č'e* is the instrumental postposition that requires an oblique complement), and the notional object is in the oblique form:

(i)   Adyghe
        cᵤəweč'ə-m-č'e   cᵤəwe-r   cᵤə-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 et al. 2009: 54, glosses
        modified from the original)

Path type, and Germanic languages instantiate the Manner type. Compare the contrast between Spanish and English in the following example:

(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 clearly 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 the example below:

(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: 57). To illustrate:[17]

---

[17]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 2009). However, the distribution of *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).

(11) Adyghe

    a. č'ale-m   ʁwəč̣ə-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   ʁwəč̣ə-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: 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: 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 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: 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.

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: 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 nonstandard, uneducated variety of Georgian (see also Chapter 12 for some discussion of this issue).

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 Peter (Pëtr) von Uslar laid the foundations of modern study of Caucasian languages for the three indigenous families. They were not linguists 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 1903, 1905, 1908, 1928a, b; Uslar 1887, 1888, 1889, 1890, 1892, 1896, 1979).[18] Baron von Uslar was also responsible for the creation of early Cyrillic-based orthographies for Nakh-Dagestanian languages.[19] The Russian-language journal "Sbornik materialov dlja opisanija mestnostej i plemjon Kavkaza"

---

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

[19] 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).

(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). Several outstanding Russian linguists worked in the area in the 1930-60s, 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 (Kibrik & Kodzasov 1988, 1990; Kibrik et al. 1977a, b; Kibrik 1977a, b, c; Kibrik 1982, 1996, 1999, Kibrik et al. 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 2009) is the result of one such field trip.

A good place to start for English sources is a special issue of *Lingua* edited by Helma van den Berg (van den Berg 2005) 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 (2011) 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 & 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 & Harris (1991),[20] 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 (Chumakina et al. 2007, Bond et al. 2016), nascent experimental work on languages of the area (Polinsky et al. 2012; Lau et al. 2018; Gagliardi 2012), and extensive new grammars (Forker 2013c 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, either due to dwindling populations or speakers moving away to areas where Russian or Georgian takes over (see Chapter 2; also van den Berg 1992).

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.[21] 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.

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; the chapter 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

---

[20] This volume is part of the series Greppin (1989-2004).

[21] However, Chapters 11 and 13 briefly discuss some aspects of the history of Kartvelian and Indo-European languages respectively.

largest language of the family, Georgian (for descriptions of Svan, another understudied language of the family, see Tuite 1998a and 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 mentioned above the complex nominal forms in Nakh-Dagestanian languages used to encode spatial meaning. Some researchers analyze them as postpositional phrases (see Chapter 3; Comrie and 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 1999 and Kibrik et al. 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, but 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 (se also fn. 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.[22] 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 Caucasus) and for Georgian. Appendix II serves to show the most common correspondences between Cyrillic, Latin, and 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

---

[22] 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*.

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 often 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 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