1 The Balkan Peninsula and Its Languages

1.0   Introduction

In this chapter, we set the stage for talking about the Balkan languages by giving relevant details about the geography, history, peoples, languages, and writing systems of the area.

1.1   Geography

Physical objects interpretable as boundaries occur in nature – mountains and rivers are the most obvious examples – and these natural boundaries can influence human history and, as a part of that, linguistic developments, cf., e.g., the role of the Rhine and the Danube in defining the borders of the Roman Empire and the spread of Romance or of mountains and valleys in defining the limits of languages and political entities of Daghestan. At the same time, however, human movement is such that natural barriers can always be crossed, as shown by the Roman occupation of Dacia and subsequent development of Romanian or the presence of Nakh-Daghestanian languages such as Tsez in Georgia, south of the Caucasus Mountains. Moreover, one group's boundary can serve as another group's link, as in the case of the Upper Kolpa River Valley, where the dialects on either side of the river formed a single group until the river became part of the administrative and then international boundary between Slovenia and Croatia (Knežević-Hočevar 2009). It is thus the case that the definition of almost any geographic region, regardless of the natural features chosen, will be, to some extent, arbitrary or else politically or ideologically determined.[1] A striking example of such determination is to be found in *Webster's Geographical Dictionary*, where the border between Europe and Asia that falls between the Black and Caspian Seas

---

1 We are using *ideology* here in the anthropological senses of Friedrich 1989 and Silverstein 1979, cf. also Friedman 1997a, 2017b; Gal & Irvine 1995; Herzfeld 1982; Woolard 1992; Woolard & Schieffelin 1994; Schieffelin, Woolard, & Kroskrity 1998; Kroskrity 2000b; Irvine & Gal 2000; Fielder 2018; and Gal & Irvine 2019. Also see Darques 2017 for an important discussion of boundaries in the Balkans.

is defined as the Caucasian Ridge under the entry for Europe, but as the Turco-Iranian political border under the entry for Asia (Bethel 1949: 347, 74).[2]

The use of a term meaning 'Balkan peninsula' as a geographic designation for Southeastern Europe is first attested in German in 1808 (Sundhaussen 1999: 31, Todorova 1997: 25) and in English in 1827 (Todorova 1997: 25).[3] The use of *Balkan* as a political term to replace the increasingly inapplicable *Turkey in Europe* – a designation that during the course of the nineteenth century had shrunk from including all of eastern Europe south of what is today Croatia, Vojvodina, Transylvania, Bucovina, and Ukraine to a strip of territory comprising the modern day countries, provinces, and regions of Albania, Epirus, the Sandžak (of Novi Pazar), Kosovo, southern Serbia, Macedonia, and Thrace – appears to date only from 1886, when J. G. C. Minchin in *The Growth of Freedom in the Balkan Peninsula* writes: "The Russians dismembered the Ottoman Empire, while she succeeded in keeping the Balkan States weak and divided" (OED s.v. *Balkan*). Derogatory usage connected with *Balkan* appears to date only from the early twentieth century. Todorova 1997: 33–37 cites a passage from the December 20, 1918 *New York Times* in which *Balkanization* is used to mean 'devastation' or 'ruin,' and she correctly points out that the meaning of political fragmentation emerges not during the period when the Balkan states were breaking off from the Ottoman Empire but rather at the end of World War One with the establishment of nation-states in Eastern Europe taken from territory that had been ruled by Austria-Hungary, Germany, and Russia: "Great Britain has been accused by French observers of pursuing a policy aimed at the Balkanisation of the Baltic provinces" (1920, OED s.v. *Balkanization*). It is ironic that the term Balkanism, when used geopolitically, means fragmentation due to conflict (cf. Todorova 1994) while its use as a linguistic term means precisely the opposite, i.e., a shared feature due to linguistic contact, in other words, interpenetrating coexistence.[4]

---

2 Cf. also Dako 1919, which refers to Albania as being in the Near East.

3 *Balkan* is the Turkish word for 'forested mountain' and also designates a mountain chain running through today's central Bulgaria known in Bulgarian as *Stara Planina* 'old mountain' (see also Račeva 1996), and in Greek and Latin as *Haimos* and *Haemus*, respectively. The Greek is of obscure etymology but may be a derivative of a word αἱμός glossed in Hesychius with δρῡμός 'copse, thicket,' a reasonable characterization of the typical Balkan terrain and in keeping with the Turkish designation; the Latin is a borrowing from the Greek term. Sundhaussen 1999: 33–34 proposes a differentiation of *Southeastern Europe* and *Balkan* in terms of physical geography for the former and human geography (history, culture, ethnography, etc.) for the latter.

4 It is arguable, however, that both usages have their origins in Western European (Great Power) constructions of the Balkans as *Other*. In the case of political fragmentation, the usage has its origins in countries whose model of political organization at the time was Empire (cf. Bakić-Hayden 1995 on British ideological sympathy for the Ottoman Empire). In the case of linguistic convergence, the usage dates from the period when the intellectual thrust of linguistic research was aimed at describing and explaining differentiation; the so-called genetic or genealogical historical linguistic paradigm was dominant, and then-current language ideology described language contact phenomena in terms of "pollution" or "corruption" (e.g., Schleicher 1850: 143, cf. Friedman 1997a). Thus, it could be argued that this irony of opposites has its origins in the same overarching nineteenth-century Western European ideology. On projects of purification in the construction of modernity, see Bauman & Briggs 2003.

As a geographic entity, the Balkan peninsula is relatively unproblematically defined on three sides as the land mass bounded by the Black Sea, the Sea of Marmara, and the Aegean to the east, the Mediterranean to the south, and the Ionian and Adriatic Seas to the west. In modern political terms, the Balkans – or Southeast Europe – are most frequently understood as comprising Albania, Bulgaria, Greece, Romania, Turkey in Europe, and the republics that made up former Yugoslavia. This definition, however, is not uncontested.[5] In addition to the problems mentioned in footnote 5 concerning Croatia, Slovenia, and Greece, there is the fact that the boundaries of Romania have at times included the territory between the rivers Prut and Dniester (Bessarabia, now the Republic of Moldova) and at others excluded Transylvania and Bucovina (see Jelavich 1983: 288).[6] A historical-political definition combining the maximal extent of the Ottoman Empire and Hungarian Crown lands would exclude most of Slovenia but include Slovakia and Hungary (Hinrichs 1999a: 34).

The northern boundary of the Balkans is especially ideologically fraught and politically implicated. Thus, in Vienna, the Balkans begin south of Vienna; in Ljubljana, south of Ljubljana; in Zagreb, south of the river Sava; elsewhere in Croatia, south of the Una; along the Dalmatian coast, it is the Dinaric Alps and not the Adriatic Sea that is taken as the boundary, etc. This is a kind of *recursiveness,* a semiotic process defined by Gal & Irvine 1995: 974 as a projection of a distinction salient at one level onto another level. In this case, the distinction is one between the desirable, progressive, Occidental, European on the one hand and the undesirable, backward, Oriental, Balkan (Turkish), on the other.[7] Moreover, a geographic

5 In fact, the northern and southern boundaries both become problematic insofar as the term *Balkan* is interpreted not as a geographic but as a cultural or historical term, one which is rejected by some Greeks and Croatians (and Slovenes). This rejection was implicitly supported by the 1962 *National Geographic* map of the Balkans (February 1962, Atlas Plate 39), whose southern boundary ends just above the thirty-ninth parallel. This rejection of a "Balkan" unity internalizes negative Western views of the Balkans (cf., e.g., Todorova 1994; Bakić-Hayden 1995) as seen in the following news item: "Tudjman got hearty applause when he said: 'Reintegration of Croatia into the Balkans is totally unacceptable for the Croatian people . . . Croatia belongs to Central European and Mediterranean circles. A short Balkan episode in Croatian history [i.e., its inclusion in Yugoslavia] must never be repeated . . . We should add a new article, a constitutional ban on attempts to merge Croatia with any Yugoslav or Balkan state or federation.' Tudjman said Croatia would enter into agreements with Balkan countries only when it was a member of the EU and could act together with its EU partners" (Patrick Moore, *Open Media Research Institute Daily Digest*, No. 16, Part II, January 23, 1997). While it is true that in linguistic terms Slovene and Croatian dialects were never counted as part of the Balkan sprachbund (see below), it is worth noting, in the context of the shifting ideology of geographic definitions, that in 1922, after the territorial divisions following World War One, the *Encyclopedia Britannica* (Vol. 30, p. 370) referred to Istria and parts of modern Slovenia as *Balkanic Italy.* This was apparently connected with the association of all South Slavs with the Balkans, since the territories of Balkanic Italy were never under Ottoman rule and are north of the forty-fifth parallel.

6 Other vagaries in the boundaries of the Romanian state have involved conflicting claims with Bulgaria over Dobrudja (Rmn Dobrogea). These claims, however, are easily viewed as Balkan-internal and therefore do not affect attempts at defining the limits of the application of the term *Balkan*.

7 We use the term *Occidental European* as a cover term for various desirable geopolitical and socioeconomic groupings such as *Mediterranean* (as opposed to *Levantine*) or *Central Europe* (as opposed to *Eastern Europe*).

boundary cannot be set in any nonarbitrary way that is applicable without qualifications to Balkan linguistics. A northern boundary starting at the Julian Alps or the rivers Drava, Sava, Kupa, or Una (the latter two being tributaries of the Sava) and then following the Danube will all exclude various parts of former Yugoslavia and all of Romania, except Dobrudja, while a line following the forty-sixth parallel roughly from the gulf of Trieste to the mouth of the Dniester would leave out northern Romania, Moldova, and Slovenia. A definition appealing to the Julian Alps and the Carpathian Mountains omits Transylvania and must traverse the Pannonian plain and Ukrainian steppe in an arbitrary fashion. Newbigin 1915: 15–16 identifies the northern boundary of the Balkans as a line running along the Lower Danube, Sava, Kulpa, and then on to Rijeka (Itl Fiume). He goes on to write: "How artificial this 'geographical' frontier is may be realized from the fact that only along part of its course does it correspond to political boundaries." Note here also the assumption that somehow the political should correspond to the "natural."

The linguistic Balkans, i.e., the Balkan sprachbund, do not correspond to any version either of the geographic or of the political Balkans. Although clearly possessing at least some antecedents in the ancient and medieval periods (and conceivably even earlier), the crucial developments that produced the current shape of the Balkan sprachbund occurred in the politico-historical context of the Ottoman Empire, whose boundaries varied considerably from the fourteenth to the twentieth centuries.[8] At no time, though, did political boundaries coincide exactly with what could be defined as the linguistic area, in part because the languages of interest are overlapping or comprise continua and do not occupy discrete territories, just as in multilingual situations languages are not limited to discrete individuals in a one-to-one relationship.

Thus, if we take the contiguity of dialects as a defining factor, we encounter difficulties in identifying a distinct geographic region that coincides with areal features. The modern line demarcating contiguous Albanian dialects, beginning at the Adriatic, includes the southernmost portion of Montenegro, all of Kosovo but the northernmost district (Leposavić), and three adjacent districts in Serbia (Medvedje, Bujanovac, and Preševo (Alb Presheva)), after which the boundary heads south into Macedonia, although as recently as a century ago, Albanian dialects extended north into the Sandžak (of Novi Pazar), northeast to the Morava at Aleksinac, and eastward beyond Vranje and Kratovo (Sax 1878). Still, taking this current distribution as our reference point, if we look at contiguous Slavic dialects, they overlap or intersect the Albanian dialects all along their northern border and continue north as far as Italy and Austria.[9] Defining the northern border of Balkan Slavic is not a simple matter. Sandfeld 1930: 1–2 uses both Serbian and Serbo-Croatian without any further qualification, Schaller 1975: 35 classes Serbo-Croatian with Greek as *zweiten Grades* ('second degree') Balkan

---

8 See also Asenova 2002: 214 which has a table giving the comparative chronology of Balkan future formation that clearly illustrates the temporal orientation defined here. Heřman 1968 offers morphological simplification in nominal systems as a Balkan linguistic boundary-defining feature.

9 We are leaving to one side the displacements caused by wars in the post-1991 period.

languages, Feuillet 1986: 37 labels Serbo-Croatian with Turkish as *peripheral* and excludes them, and Asenova 2002: 16 observes that Greek is a part of the Balkan sprachbund, whereas Serbo-Croatian is not. Hinrichs 1999a: 339–428 includes separate chapters on Serbian, Croatian, and Bosnian, but these language outlines make no attempt to address questions of the Balkan sprachbund. In these matters, we agree with Asenova 2002 that Greek is definitely part of the Balkan sprachbund. Although it is the case that certain "classic" features are absent from Greek or represented in some different sense (the primary example being the postposed definite article, on the one hand, but morphological definiteness, on the other), the fact remains that Greek does participate in the majority of key features, including some "partial" ones (i.e., bilateral and trilateral correspondences); see now also Joseph 2020b. It is demonstrable, however, that those features most characteristic of the Balkan sprachbund found in the former Serbo-Croatian are represented in the greatest number in the Southern Bosnian/Croatian/Montenegrin/Serbian (BCMS) dialects, also known as the Torlak dialects and also still claimed as Bulgarian by many Bulgarian linguists (e.g., Kočev 2001: 55; Tetovska-Troeva 2016, but cf. Friedman 2006e, 2021; Sobolev 2020).[10] The northern boundary of these dialects varies according to whether they are claimed as BCMS on the basis of phonological criteria (e.g., the absence of phonological length and tone, Ivić 1985) or Bulgarian on the basis of morphological criteria (e.g., the presence of a postposed definite article (S. Mladenov 1929; Kočev 2001: 55)), but at its northernmost, the demarcation begins in Kosovo at the Albanian border just south of Dečani (Alb Deçan) and goes to Obilić (Alb Kastriot) at the confluence of the Lab and Sitnica rivers, then follows the Lab northeast, turns east to the south of Podujevo, north to Stalać (west of Prokuplje) then east across Mount Rtanj, south of the River Crna and continuing to the Bulgarian border south of Zaječar (Ivić 1985). In the case of Romanian and Moldovan, we take the current political borders for the sake of convenience. As Sandfeld 1930: 2 points out, the close connection between Balkan Romance north and south of the Danube (and, we might add, between Bulgarian and Romanian in terms of historical interaction) necessitates its inclusion.

---

10 In general, we use the term *Serbo-Croatian* when that is the term used by a cited author or when the reference is to the time period 1850–1991. (While it is true that the Vienna Literary Agreement of 1850 did not use the term *Serbo-Croatian*, it is still viewed by many as the foundational moment of that literary language.) The abbreviation BCMS (Bosnian/Croatian/Montenegrin/Serbian) refers to the former Serbo-Croatian or the dialect complexes covered by that term both for the post-1991 period and when no specific temporal location is relevant. Owing to the ethnic basis of the current BCMS situation, the same dialect can be claimed as Bosnian, Croatian, or Serbian depending on the religion of the speakers. This includes the Torlak dialects of southern Serbia and Kosovo (cf. Lisac 2003: 143–153 and Ivić 1958, who uses *Torlak* as a cover term for the Timok-South Morava-Prizren dialects). See Chapter 2 for a general overview of sprachbund features. On the Torlak dialects, see especially Alexander 1981, 1983, 1984–1985, 1993, 1994. See Sobolev 1998a for an attempt to distinguish a border between eastern Serbian and western Bulgarian (also Sobolev 2020, which argues for the uniqueness of Torlak). Birnbaum 1965: 39–40 summarizes earlier debates over the relative "Balkanness" of BCMS dialects. We should note here that some southernmost Montenegrin (Zeta-Lovćen) dialects also show some significant Balkan features as a result of contact with Albanian (Stojanovič 1935; Miletić 1940; Pešikan 1965; R. Greenberg 2000; Morozova & Rusakov 2018ab; Morozova 2021).

Owing to their relatively contiguous extension over the Balkan peninsula, Slavic dialects, combined with the extent of Albanian and certain features in Romani (e.g., the analytic future in *ka*, see Boretzky & Igla 2004: 1.172–174, 2004: 2.244) provide a measure of boundary definition that differs from what can be deduced from the other language groups. In this sense, Leake 1814 was not entirely mistaken in the position he assigned to Slavic in Balkan linguistic contact, for it is precisely on current South Slavic and adjacent territory that features spread and diminish.

For the remainder of this chapter, we take elements from the geographic, political, historical, and linguistic definitions of the Balkans and limit ourselves to the territories of modern (European) Turkey, Bulgaria, Greece, the Republic of North Macedonia, Albania, southern Serbia, Kosovo, and Romania. Albanian-Slavic contacts in Montenegro and the Sandžak as well as Balkan languages in diaspora are beyond our focus as is the situation in Bosnia-Hercegovina, Croatia, Slovenia, and northern Serbia (Šumadija, Vojvodina, etc.), although, where relevant, references to some of these places and their languages are made. We acknowledge the ideological underpinnings of the construction of the Balkans as a geographic entity and the cline-like nature of the linguistic convergence area.[11] That said, however, those constructions have their bases in certain concrete social and linguistic phenomena, and those phenomena are the focus of this book (see Bakić-Hayden 1995; Todorova 1994, 1997; Friedman 1997a; for a specifically British point of view see Goldsworthy 1998).

1.2   Languages

Having addressed the issue of defining the Balkans as a geopolitical space, we can now turn to the languages spoken there and those among them constitutive of the linguistic league that is the subject of this work. As explained in the introduction, there are four living Indo-European language groups that are universally recognized as containing members that constitute the classic Balkan sprachbund – Balkan Albanic, Balkan Hellenic, Balkan Romance, and Balkan Slavic – to which we added two other groups, one Indo-European, Balkan Indic, and one not, Balkan Turkic.[12] The relevant languages in these groups – and the concept of *language* itself – are treated below. We begin, however, with two groups of languages that are spoken or were spoken in the Balkans but which are not

---

11 As Masica 2001: 241 perceptively notes regarding the correlation between verb-final word order and the use of postpositions in South Asia: "What appears as an interconnected bundle at the core of these areas comes apart – 'frays', as it were – at their margins." In the case of Balkan Slavic, the presence of both contact-induced innovations and contact-maintained archaisms gets progressively weaker as one moves north and west through BCMS territory (Friedman 2003c), and while we comment on those "frayed" margins where appropriate, our focus remains on the core.

12 The term *Albanic*, unlike the other glossonyms in *-ic*, is not well established, but we introduce it here as a convenient and terminologically consistent label for all modern speech forms identifiable as Albanian as well as whatever the ancient Indo-European ancestor of Albanian was.

scrutinized here, namely dead languages of the Balkans with no significant attestations above the level of the lexicon – if even that much – and languages which, while spoken in the Balkans, do not enter into the Balkan sprachbund. These two groups can be classed together as *languages of the Balkans* in opposition to *Balkan languages*. By *languages of the Balkans* we mean those languages that are or were spoken in the Balkans but do not display significantly (or in any attested fashion) the morphosyntactic and other convergence phenomena that are central to our concept of a contact area (sprachbund), while by *Balkan languages* we mean those languages that do. Within the language groups that contain Balkan languages, there are also non-Balkan languages and what we can call extra-Balkan languages. The non-Balkan languages, as indicated in the introduction, are those which were never spoken at all or by any significant population in the Balkans, while the extra-Balkan languages (or dialects) are those that have emigrated out of the Balkan linguistic contact area.

1.2.1 Dead Languages of the Balkans

While potentially of considerable significance for the study of Balkan linguistics, the poorly attested dead languages of the Balkans – given the present state of our knowledge – cannot tell us anything about the morphosyntactic and other convergences observable in the modern Balkan languages and thus do not present any explanatory evidence for observed later structures. This is made clear in Woodard 2004: 9–15, the most thorough and up-to-date compendium of the state of our knowledge of the world's ancient languages, in which most of the ancient Balkan languages were judged too meagerly attested to be included in the work's grammatical descriptions.

The single potential exception to the abovementioned grammatical lacuna was analyzed by Hamp 1982: 79, who concludes after careful etymological argument that the name of the ancient site of *Drobeta* – located on the Danube near modern Turnu Severin in northwestern Oltenia (Romania) – contains "a Latin misunderstanding or misparsing in Moesia Inferior of *druu̯ā–tā*, a definite noun phrase with postposed article." As such, it gives "direct evidence in the Roman period of one of the most notable syntactic constructions of the Balkan sprachbund, i.e., a specimen from the autochthonous language of the model of the Romanian postposed article which was calqued out of Latin materials." Moreover, it constitutes "direct attestation for the common possession of this important feature linking modern Albanian with Moesia Inferior" (cf. also Hamp 1990). Note that Hamp uses the phrase *autochthonous language*, thereby avoiding any of the linguistic names that we actually possess. While Hamp 1981–1982, 1992a has also shown masterfully how etymologies of toponyms such as *Ohrid*, *Kukës*, and *Prizren* can demonstrate adaptation from one language to another thus revealing social, historical, and linguistic data that cannot be otherwise found in texts, it nevertheless remains the case that these products of language contact do not tell us anything about modern morphosyntactic convergences. From this it follows that, aside from the highly

suggestive datum adduced in Hamp 1982, all appeals to linguistic substrata as explanations for morphosyntactic convergences (e.g., Kopitar 1829; Miklosich 1862;[13] Solta 1980: 210, 223) do not rest on any concrete evidence whatsoever. It is therefore the case that while such linguistic substrata may indeed be responsible for subsequently observed phenomena, any speculation beyond that which we have just cited has no basis in existing evidence.

Nevertheless, we survey here the principal dead languages of the Balkans and our state of knowledge about them, with references for further research for the interested reader.

1.2.1.1 Ancient Macedonian

Unlike modern Macedonian, which is a Slavic language, Ancient Macedonian – the language of the core troops of Alexander the Great (Katičić 1976: 106) – was an Indo-European language of uncertain connections within Indo-European. Katičić 1976: 108–112 identifies three layers in Macedonian vocabulary. One layer is clearly Greek and likely borrowed. Another layer has no Indo-European connection, and a third layer is clearly related to Greek but the words appear to be cognates rather than borrowings. The question, thus, is this: (1) Was Ancient Macedonian a Hellenic dialect distinct from the Hellenic dialect that gave rise to Ancient Greek (with all of its dialect diversity, on which see, e.g., Buck 1955), one whose non-Greek-looking vocabulary consists of loanwords, Greek words that did not survive in our known Greek sources (cf. Katičić 1976: 111), and reflections of inadequacies in the Greek transcription of Ancient Macedonian pronunciation, or (2) was it a sibling to Hellenic, i.e., part of a single pre-Hellenic diasystem that subsequently branched into Hellenic (whence the attested Greek dialects) and Ancient Macedonian, or (3) was it a non-Hellenic Indo-European language, possibly allied to one or more other dead and poorly attested languages of the Balkans?[14] Woodard 2004: 14 adduces the example of Ancient Macedonian *kebalá*, and an apparently related form *gabalá* glossed as 'head' by Hesychius without identifying the language, and Greek κεφαλή 'head.' This and other glosses (e.g., AMac *abroutes* ~ AGrk ὀφρύες 'eyebrows,' from PIE *H_3**bh**ru-, cf. Skt *bhrū-*, Eng *brow*) suggest that Ancient Macedonian retained the voicing of Indo-European voiced aspirates (e.g., *bh*), but, perhaps, lost the aspiration, while Greek (or Hellenic) kept the aspiration but not the voicing. The question is unduly complicated by modern nationalist attitudes, according to which a determination of the provenance of

---

13 This work, volume 12 in its series, is often cited as Miklosich 1861, even though it appeared in 1862, because it is typically bound together with volume 11 from 1861. We have decided to cite it as 1862, on the advice of Mary Allen Johnson, curator of the Hilandar Research Library at The Ohio State University, based on the date on the title page.

14 In this section, we use *Greek* to mean the attested Ancient Greek dialects, i.e., Attic-Ionic, Aeolic, Doric, and Northwest Greek (the classification of Pamphylian and Mycenaean need not concern us here), and *Hellenic* to mean the Indo-European dialect that gave rise to Greek in all its forms as we know them. Thus, Ancient Macedonian cannot be considered a dialect of Ancient Greek on a par with Attic-Ionic, etc.

Ancient Macedonian would (somehow) justify modern-day territorial or ethnonymic claims. We can note here that Hall 1997: 62–64 makes it abundantly clear that regardless of the linguistic relation of Ancient Macedonian to Ancient Greek, the weight of evidence indicates that the Ancient Greeks did not consider the Ancient Macedonians to be ethnically Greek (see also Badian 1982), and in any case, the ancient territorial and ethnonymic situation is utterly irrelevant for the modern one (cf. also Ilievski 1988, 1997, 2008; Joseph 2024). Our knowledge of Ancient Macedonian comes in large part from glosses in Hesychius, a fifth-century CE glossator and antiquarian with an interest in lexical curiosities of earlier Greek and in compiling them. Various texts and dialect materials were available to him that no longer exist, and he had earlier lexicographical works to draw on too. He listed several words said to come from "Μακεδόνες," i.e., the ancient Macedonians. In addition, there are several place names and personal names that are either known from historical sources to be Ancient Macedonian or are presumed to be so.

It is important to note that no verifiable clearly Ancient Macedonian inscriptions exist.[15] For details on the existing data for Ancient Macedonian and their analysis, see Ilievski 1988, 1997, 2008; Katičić 1976: 100–116; Pudić 1967; Woodard 2004: 12–14; Neroznak: 1978: 168–173; and Gindin 1987, as well as the various pro-Greek articles in Giannakis 2012, with their detailed bibliographic coverage. Evidence concerning the Epirotes is likewise scant and uncertain (Katičić 1976: 120–127).

1.2.1.2 Illyrian

It is a widely held opinion that Illyrian represents the ancestor of Albanian (cf. Katičić 1976: 184–188; Polomé 1982: 888; and Ismajli 2015 and sources cited therein). The existing data, however, are so sparse, contradictory, and/or speculative, that such a claim has been challenged. Hamp 1994a, 1994b went so far as to question whether the term *Illyrian*, even in the narrow sense, actually refers to a single language or was rather a cover term used by ancient authors for a group of languages spoken by various Indo-European tribes that may or may not have been more closely related. (Cf. popular usage of the term *Australian Aborigine*, e.g., in the BBC television series "Dr. Who" ["Four to Doomsday," Episode 1, first aired January 18, 1982], as if it referred to a single language rather than hundreds of different languages.)[16] As Woodard 2004: 11 notes, we do not possess a single

---

15 The Pella Curse Tablet, discovered in 1986 in a part of the territory of Ancient Macedonia currently located in modern Greece and dating from the fourth century BCE, is written in a distinct form of Doric (Northwest) Ancient Greek. However, this does not prove that Ancient Macedonian was an Ancient Greek dialect. It only proves that some form of Ancient Greek was in use in Ancient Macedonia as a written language and was probably spoken there by some people whose status in Macedonian society we do not actually know. Moreover, it dates from the century when Ancient Macedonian ceased to be spoken, the inhabitants having shifted to Greek. See Méndez Dosuna 2012 for an attempt to argue that the obvious differences between Ancient Greek and Ancient Macedonian are due to inadequacies in Greek orthography. Joseph 2024 offers a discussion of the ideological angles to the interpretation of the Ancient Macedonian material, with a survey of relevant literature.

16 See Katičić 1976: 165–177 on the multiplicity of definitions of Illyrian.

verifiable inscription in Illyrian. Most of the evidence for Illyrian that is adduced is based either on onomastics – which, in addition to being highly speculative (we cannot be sure that an element found in a proper name actually corresponds to a lexical item that it happens to resemble in some other language), clearly does not form a single consistent whole (Polomé 1982: 866; Sh. Demiraj 2004: 55–56; cf. also Katičić 1976: 154–166, Tzitzilis 2007a: 746) – or on the problematic assumption that Messapic, a poorly attested and inadequately deciphered Indo-European language of southern Italy, was a relative or dialect of Illyrian (see Polomé 1982: 866; Hamp 1994b).[17] As Woodard 2004: 15 observes, however, until and unless Illyrian is better attested, a connection with Messapic must remain an unverifiable hypothesis (but see now Hamp 2008).[18] Our total unambiguous set of attested

17 This latter assumption is based on classical sources indicating a connection between Illyrian and Messapic, e.g., Pliny 3.11.102: *Brundisio conterminus Paediculorum ager; novum adulescentes totidemque virgines ab Illyriis XII populos genuere* 'Adjacent to Brindisi is the territory of the Paediculi, whose twelve tribes are the descendants of nine youths and maidens from the Illyrians' (cf. Katičić 1976: 163; Polomé 1982: 866; Woodard 2004: 15). In the absence of additional consistent supporting evidence, we are left with tantalizing speculations but no certainties. In this regard Katičić 1976: 169–170 cites the instructive history of an inscribed ring excavated from a grave in Kalaja Dalmaçes near Shkodër in northern Albania with the clearly legible inscription ANA OHΘH ICER. It was assumed to be Illyrian and interpreted on the basis of Messapic. In Messapic inscriptions, *ana* occurs as a title used before the names of goddesses and *isareti* as a plausibly interpreted verb 'consecrates' (Indo-European *isH_1ro- 'sacred,' cf. Grk ἱερός, Skt *iṣira*-) leaving OHΘH as the name of the goddess. Upon further investigation, however, it turned out that the grave and its ring were Byzantine and the inscription, when read bottom to top, was Medieval Greek: K[YPI]E (with the IC characters taken as a < K >) [B]OHΘH (for BOHΘEI) ANA (for ANNA), i.e., 'Lord, help Anna.'

18 Eric Hamp 2008 determined on the basis of the parallel between Albanian *flê-fli* 'sleep.3SG-IMPV.SG' and Messapic *fli* 'sleep' that there was indeed an ancient relationship between Albanian's ancestor and Messapic. The /i/ represents the outcome of Winter's Law – lengthening before a PIE plain voiced stop (a *media*, cf. *rronj* 'endure' < *$r\bar{e}g$-n (with o as the regular outcome of PIE *$\bar{e}$; the root is that of Greek ὀρέγω 'reach, stretch'). The root here is *leg- 'lie' (with f- representing an ancient preverb). The lengthening represents an innovation that Albanian shares with Balto-Slavic (cf. Slv *ĕd*- 'eat' < *$\bar{e}d$-, the root being that of Grk ἔδω 'eat') from the period before the speakers of the Indo-European dialect that became Albanian (and Messapic, and Illyrian) left northern Europe; the nasal infix in the root (seen in the Geg nasal vowel) is another trait shared with Balto-Slavic. Moreover, Hamp argues that the possibly Illyrian *sybina* 'hunting spear' (found also in AGrk as σιβύνη with a variant συβίνη) can be compared with Geg Albanian *thupën* (standard *thupër*) 'withe, cleaning rod' and that Latin *sīca* 'dagger' – a word of obscure etymology that is likely a borrowing – is cognate with Albanian *thikë* 'knife.' These two words (*sybina* and *sīca*) are cited together as Illyrian or connected with Illyrians in Ennius, Annales (249; Paulus ex Festo 500, 10): *'Sybinam' appellant Illyrii telum venabuli similae. Ennius – Illyrii restant sicis sibinisque foedentes.* '"Sybina," a name given by the Illyrians to a javelin resembling a hunting spear. Ennius – The Illyrians stood fast and stabbed with curving knives and hunting spears.' Caution is needed, to be sure, as apparently related forms meaning 'javelin' are found in Armenian (*səvin*) and Syriac (*swbyn*) and possibly related forms with -g- occur too, e.g., AGrk σιγύν(ν)ης 'hunting spear,' reminiscent of the name of an apparent Iranian tribe Σιγύνναι. So these various forms may reflect an old Wanderwort, thus introducing additional complexities to the possible origin of *thupën* in relation to *sybina*. See Beekes 2010: Vol. 2, 1327–1328 for some consideration of the range of forms here. In Hamp's latest thinking (p.c., 2013), speakers of the dialects that became Messapic, Illyrian, and what Hamp calls Albanoid divided at what is today's Gulf of Trieste. One group traveled down Italy to become the Messapians, another traveled down the Dalmatian coast to become the Illyrians (or one of the groups covered by the term *Illyrian*, since our data are so scanty and the term was employed variously by ancient authors), while a third group, comprising speakers of Albanoid, was more to the east.

lexical items that are labeled by ancient authors as Illyrian consists of three words – *rinos* 'mist,' *sabaia* 'fermented fruit beverage,' *Deuadaia* 'satyr' (Katičić 1976: 170–171; Tzitzilis 2007a: 746) – and a possible fourth *sybina* or *sygine* 'hunting spear' (Duridanov 1999: 755; Polomé 1982: 867).[19] The first of these is plausibly connected with Alb *rê* (older *ren*) 'cloud,' the second seems cognate with Eng *sap* (Skt *sabar*), the third may be connected to Grk θύω 'rage,' and, if the fourth is admitted, a cognate with Armenian *suin* [səvin] 'spear' has been suggested, but this is hardly adequate proof that we are dealing with a single language, much less one whose modern descendent can be determined with certainty.[20] Other evidence connecting Illyrian with Albanian is similarly flawed. Thus, for example, the toponym *Dalmatia* has been associated with Alb *dele* 'ewe,' with variant form *delmë* (Meyer 1891: s.v.), supported by evidence from Strabo (7.5.5) referring to *Delminium*, the capital of the Delmatae, as πεδίον μηλόβοτον 'a pasturage for sheep' (cited in Sandfeld 1930: 143). In fact, however, the passage in Strabo does not say that Delminium was a pasturage for sheep, but rather that Nasica turned it into one, i.e., he razed it – the expression is an idiom for utter destruction and cannot be interpreted as a literal reference to sheep (Katičić 1976: 173). The resemblance of the ethno-toponym *Dardania* to Albanian *dardhë* 'pear' (cited in Sandfeld 1930: 143) is suggestive, but nothing more. As with the Ancient Macedonians, so too with the Illyrians, at times modern nationalist claims of legitimacy based on autochthony or illegitimacy based on migration interfere with objective evaluations of the available evidence; and as with the Ancient Macedonians, so too with the Illyrians, such considerations have no place in modern linguistics nor should they be taken as the basis for modern geopolitics. Regardless of what the ancient situation may have been, the modern one is what we have, and the question of Illyrian remains open, even if some sort of connection with Albanian seems reasonable. In any case, as with the other dead languages of the Balkans, we do not have a single shred of unequivocal textual evidence above the level of the lexical item.

1.2.1.3 Thracian

We have considerably more lexical evidence for Thracian than for either Ancient Macedonian or Illyrian, e.g., *brûtos* 'beer' (cf. OE *breowan* 'to brew' (Woodard 2004: 12)), and classical sources tell us that it was widely spoken in the region that we call the Balkans today. According to Strabo, the Getae shared their language with the Thracians and the Dacians spoke the same language as the Getae (Katičić 1976: 131–132). We know that Ovid wrote poetry in the Getic dialect, but none of it has survived (Katičić 1976: 137). We have two short inscriptions in Thracian, but, as Katičić 1976: 137 observes, they "are much too scant and their interpretation not

---

19 But see also footnote 18.
20 Neroznak 1978: 162 gives a number of additional glosses, most of them Epirotic, and observes that we cannot know if these are even from the same language as the glosses identified as Illyrian.

sure enough to be of any real value for the study of Thracian."[21] We also have what appears to be a Dacian inscription – *Decebalus per Scorilo* which Georgiev 1977: 199 glosses 'Decebalus son of Scoril[o]' (cf. Duridanov 1999: 749). However, as Brixhe 1994: 186 points out, the inscription is in running script and thus the segmentation is far from certain. Georgiev 1977: 193–215 has adduced evidence connecting Albanian with Thracian or Daco-Moesian, and as Fine 1983: 10–11 points out: "these are serious (nonchauvinistic) arguments and they cannot simply be dismissed." Nonetheless, in the absence of more extensive, genuinely textual, evidence, we cannot be certain what, if any, the relationship was of Thracian (or Dacian/Daco-Moesian) to Albanian. Fine 1983: 11 further writes: "More evidence is needed which, owing to the nature of our sources, may never be obtained; thus the question may well be one of many in early Balkan history which we may never be able to answer." The possibility of a relationship between Thracian and Phrygian is likewise uncertain and vexed (Woodard 2004: 12; Brixhe 2004: 780).

1.2.1.4 Dacian (Daco-Moesian)

Dacian[22] is identified as the Indo-European language spoken on that part of the territory of modern Romania that was occupied by the Romans starting with the first battles in 101 CE to the organized migration south of the Danube in 271 CE. Moesian has been identified as a language or dialect south of the Danube closely allied with Dacian (Georgiev 1977: 193–211). While modern Romanian is a Romance language, it is generally assumed that remnants of the Dacian lexicon persist, and moreover the existence of a group of about seventy to ninety Indo-European cognates shared between Albanian and Romanian, e.g., *mal* Alb 'mountain' – Rmn 'river bank,' Alb *sorrë* – Rmn *cioară* 'blackbird' (from PIE *$k^w\bar{e}rsn\bar{a}$ 'the black (thing)'), Alb *moshë* 'age' – Rmn *moş* 'old man,' is taken as indicative of a close prehistoric connection between the ancestor of Albanian and the pre-Latin language of the territory that is included in modern Romania.[23] Given that Dacian (or Daco-Moesian) appears to have been closely allied to Thracian and that the evidence that is taken as constitutive of Illyrian appears to belong to a different Indo-European group, the question naturally arises whether the evidence of the Albanian–Romanian connection points to a genetic/genealogical or areal link. Hamp 1994a is quite clear in suggesting that Romanian and Albanian appear to have shared in intense Romanization, with the language that became Romanian having been more or less fully Romanized while that which became Albanian having been on the road to such a shift but not having undergone it. But, as noted above, Hamp is extremely cautious regarding nomenclature and suggestions of

---

21 See Duridanov 1999: 737–739 and Panayotou 2007 for details, but see also Brixhe & Panayotou 1994.
22 On Moesian, see also Papazoglu 1978: 402–443. Woodard 2004: 12 identifies Moesian as that form of Thracian spoken north of the Balkan (Haemus) range.
23 Various lists of these forms with differing degrees of discussion are available, e.g., Polak 1958; Sh. Demiraj 2004: 172–173; Sala 1999: 151–166; Asenova 2002: 47; see also §4.2.1.1.

genetic versus areal relationship in this regard, and we share in his caution.[24] Hamp 1994b, 1999 uses the term *Albanoid* to denote the stage of Albanian after it diverged from the Balto-Slavo-Albanian dialect of late Proto-Indo-European, but before it came into contact with Latin and the Roman empire (circa 200 BCE). *Proto-Albanian* refers to the period when the language was in contact with Latin, but before contact with the Slavs (circa the seventh century CE). *Common Albanian* refers to the variety of Albanian that all extant varieties must have descended from.

1.2.1.5 Pre-Greek, Pre-Hellenic, or "Pelasgian"

There are numerous words that are part of the Ancient Greek lexicon that are clearly not Hellenic, that is, not direct inheritances from Proto-Indo-European. While some of these are borrowings from identifiable sources, e.g., χιτών 'tunic' from a Semitic language (cf. Phoenician *ktn* 'linen tunic') or κύανος 'copper' from a language of Anatolia (cf. Hittite *kuwanna(n)* 'copper'), many do not have a recognizable source. Moreover, some of these words are strikingly similar to – but, crucially, not identical with – existing inherited Greek words, e.g., ἀλείφ-ω 'anoint,' which in its -λειφ- "core" seems quite similar to the root of the inherited lexeme λίπ-ος 'fat' (thus, identical but for the initial vowel and final aspirate). Furthermore, place names all over Greece have recurring apparently suffixal elements added onto stems that have no clear Greek parallels, e.g., the -νθ- found in the toponyms Κόρινθος and Ζάκυνθος, among others, and there are some words for items of material or natural culture that have similar elements, such as ὑάκινθος 'hyacinth' or ἐρέβινθος 'a kind of bean.' Putting all this together, scholars have posited, quite plausibly, at least one lexical layer, and maybe more, of words that entered Greek from an indigenous, most likely Indo-European-speaking, group that populated Greece before the coming of the Greeks early in the second millennium BCE. The language or languages of this group are usually referred to as Pre-Greek (Katičić 1976: 16–97), or Prehellenic (Hamp 1994b), and sometimes also "Pelasgian" (using an ethnonym that occurs in the Odyssey and in other ancient sources), and it is possible that other groups are to be identified in other lexical layers. The toponyms, moreover, are generally held to reflect a language of ancient Anatolia, quite possibly (but by no means assuredly) an Indo-European one, given apparent parallels in the shape of certain suffixes (e.g., -*anda*- in Anatolian place names, such as *Kuranda*, cf. Hittite *kuera*- 'field,' reminiscent of Κόρινθος). As interesting as this material is (even though much of it is speculative and based on

---

24 Cf. also Hamp 1989a: 47: "We may say that historically Romanian is Latin spoken with an Albanian stress system; i.e., Romanian reflects Danubian Late Latin to which the Proto-Albanian prevailingly penult accent was applied to morphological words and in which noun and verb phrases with clitics were constructed according to the Proto-Albanian plan of syntax and phrasal accent." Note also Hamp 1982, cited above (§1.2.1). The recognition of such a substratum effect might be useful also with regard to possible parallels in the syntax of the prepositions for 'with,' Albanian *me* and Romanian *cu*, which both, somewhat unusually in the overall Balkan context, typically take definite objects (when unmodified); see §5.5 on accent and, with important clarifications, §7.9.2 on prepositions.

similarities of form that might be accidental), and as important as it is for understanding the prehistory of the Balkans and related matters such as the coming of the Greeks into Southeastern Europe, it is entirely lexical in nature and moreover none of the languages possibly identified in this way can be shown to have contributed anything to the pan-Balkan features that emerged centuries later and gave rise to the Balkan sprachbund.[25]

1.2.1.6 Phrygian

According to ancient sources Phrygian was spoken in the Balkans during the prehistoric period (Katičić 1976: 130–131; Tzitzilis 2007b; Brixhe 2004: 777). Although the language is poorly attested, we have whole inscriptions in it, but all of them are from Anatolia (Hamp 1994b), and even in these the grammatical evidence they give us is meager (Brixhe 2004: 782–786). As a Balkan language, therefore, Phrygian is without adequate concrete textual evidence.[26]

1.2.1.7 Paeonian and Dardanian

These are ethnonyms associated with territory approximately in the center and south of today's Republic of North Macedonia and in Kosovo, respectively. The languages to which they presumably refer are so poorly attested and understood that their relationship to the other dead languages of the Balkans is a matter of numerous contradictory speculations linking them with Thracian, Dacian, and/or "Illyrian" (Solta 1980: 35–39; Sh. Demiraj 2004: 67; Katičić 1976: 117–118, 181; Duridanov 1999: 746). As Katičić 1976: 118 points out: "Paeonia is a blank on the linguistic map of the ancient Balkans." The only gloss we have is *mónapos* 'European bison' (cf. Thracian *bólinthos*, Woodard 2004: 12). Papazoglu 1978: 218–219 notes that we have three glosses in Dardanian, all of them plant names: *aloítis* 'gentiana,' *sōpîtis* 'wormwood,' and *kakalía* of uncertain meaning. While some have connected *sōpîtis* with Illyrian *sabaia*, there is no way to test the association. Moreover, even if all three plant names were to prove to be Illyrian (and Illyrian an identifiable language), we would have no way of knowing whether

---

25 For instance, one of the elements in toponyms in Greece of interest here is-*σσ*-, as in the mountain name Παρνασσός, a suffix that has a striking parallel in Anatolian place names in –*assa*-, e.g., *Petassa*. However, the Greek suffix also shows up as -ττ- in Attica (e.g., in the mountain name Ὑμηττός), and the inner-Greek -σσ-/-ττ- correspondence points to an earlier cluster with jod such as **kj*, whereas the Anatolian –*ss*- cannot reflect such a cluster. The similarity between the suffixes may thus be illusory from an historical standpoint. We draw heavily here on the material in Katičić 1976 and on discussions with H. Craig Melchert of UCLA (now emeritus) on the nature of the Anatolian evidence. See also Beekes 2014 for an attempt to marshal the pre-Greek material into a system.

26 Katičić 1976: 150–151 notes the claim in Herodotus (7.73) that the Armenians were Phrygian colonists but declines to speculate further on the matter, although elsewhere claims of the presence of prehistoric Armenians in the Balkans can be encountered (Seliščev 1925: 51; Cowgill 1960). Brixhe 1994 notes the problem that the lack of word divisions in Phrygian inscriptions renders them of very little use for morphosyntax.

or not they were loanwords. As noted in §1.2.1.2, the connection of the name Dardanian with Albanian *dardhë* 'pear' is suggestive but unsubstantiated. In the early 2000s, Kosovar Albanian nationalist practice has sought to replace the toponym *Kosovo* (a Slavic/Serbian denominal adjective from *kos* 'blackbird' understood as modifying the neuter noun *polje* 'field') with the resuscitated form *Dardania*.[27] With independence in 2008, however, the internationally recognized name of the country is Kosovo.

1.2.1.8 Celtic

The Celts were very much a presence in Southeastern Europe in ancient times, having invaded the Balkans in 279 BCE. They established settlements thereafter in what is now central Serbia, Thrace, and the Danube basin in parts of present-day Romania and Bulgaria, and they raided as far south as Delphi. Little is known about the Balkan Celtic language(s) – there may well have been more than one Celtic tribe involved – beyond personal names (see Katičić 1976: 180ff.; see also Papazoglu 1978: 50–56, 340–389, 439–491 et passim, and Sandfeld 1930: 98). However, there are a few inscriptions from the general area of ancient Noricum in southern east central Europe in what could be classified as the northern extremes of the Balkans, depending on how "Balkan" is defined (see §1.1), in particular from Ptuj (Slovenia) and Grafenstein (southern Austria). Intriguingly, the Grafenstein inscription has a sequence of letters which read *ollo=so*.[28] Some scholars have seen in this sequence 'all' (cf. Old Irish *oll* 'ample') followed by a postpositive demonstrative/article (cf. Sanskrit *sa*, and other Indo-European *s- demonstrative elements), perhaps as a noun phrase in itself or as a modifier of a subsequent word. Such a reading is strikingly reminiscent of the Balkan postposed definite article.[29] Given that the inscription is essentially uninterpretable and that this "identification" rests entirely on etymological guesswork (however enlightened it may be), it is hard to make any legitimate connection between this sequence and the better understood Balkan phenomenon. Moreover, it is not clear that this is even the same language as that used by the Celts situated more in the central and southern Balkans. Thus however tantalizing this evidence may be, it does not offer much of substance to our understanding of the ancient Balkan linguistic situation or to later sprachbund phenomena.[30]

---

27 Many Slavic toponyms in the region are based on Slv *kruša* 'pear' (e.g., *Kruševo* in North Macedonia), and *dardhë* figures prominently in Albanian ritual songs. Nonetheless, in the absence of any concrete evidence, we have no way of knowing whether the resemblance of *Dardania* to *dardhë* is coincidence or not (cf. Papazoglu 1978: 261). See Papazoglu 1978: 58–86, 124–129, respectively, on the Triballi and Autariatae, groups that were significant in the Central Balkans about which we have no reliable linguistic data.

28 A facsimile can be viewed at <www.univie.ac.at/indogermanistik> (via the "Ressourcen" link), a webpage maintained by the Institut für Indogermanistik of the University of Vienna. Other relevant online resources for this inscription include <www.arbre-celtique.com/encyclopedie/tuile-de-grafenstein-4455.htm> and Wikipedia <http://en.wikipedia.org/wiki/Noric_language>.

29 The word after *so* is *vilo*, which at present is quite obscure etymologically and contextually.

30 See also Papazoglu 1978: 345–389 on the Scordisci, the core of whom are presumed to have been Celtic, and Falileev 2009 on toponymic material.

1.2.1.9 Germanic

Germanic tribes such as the Gepids and Goths also settled in the Balkans during the late ancient period. Moreover we know that the Norse Vikings were in the Balkans since there are Runic Norse inscriptions in a few places, as described by Page 1987: 53, most notably a grafitto of "the name Hálfdanr scratched on the marble of the great church of Hagia Sophia, Istanbul, and an inscription, hardly readable these days, cut on a marble lion once at Piraeus and now outside the Arsenal, Venice." Nonetheless, their languages do not appear to have had any discernible effect, beyond toponyms and perhaps some other vocabulary, mostly for items of material culture (Mihaescu 1993: 322–323; cf. also Poruciuc 2009), but also ecclesiastical terminology that entered OCS via Frankish or OHG (Lunt 1982). Also, Jokl 1929 identified a few cases of loanwords in Albanian specifically from Gothic, such as *fang* 'sod, piece of turf' from Gothic *waggs* 'Paradise' and he suggested further that *tirk* 'trousers' is to be connected with a presumed Gothic **þiu(h)-brók-*, (lit., 'thigh-breeches'), a form safely inferrable for Gothic due to attested OHG *theobroch* (see also Barić 1954). Albanian *tirk* is of particular interest since it is related to Romanian *tureac* 'leggings' (with a by-form *tureatcă*), which is analyzed by Diculescu 1929 as also being based ultimately on this Germanic form (though he sees it as transmitted through Gepids). Although Sandfeld 1930: 97 raises the question of whether Low Latin *tubroces* could be an intermediary in the importation of this Germanic form, at which point the Romanian and Albanian words would not be direct evidence of Germanic influence in the Balkans, the analysis by Hamp 1990 has clarified how both the Romanian and the Albanian forms represent different developments out of a Germanic, and specifically Gothic, starting point.

1.2.1.10 Iranian

Iranian-speaking tribes such as the Scythians and Sarmatians extended into the Balkans and left lexical traces, e.g., in hydronyms such as *Danube*. Iranian military aristocracies are also associated with the Slavic invasions of the early Middle Ages (Fine 1983: 26, see §1.2.3.4). There are no Balkan Iranian texts, however, and so Iranian joins Germanic and Celtic as a transient linguistic group in the Balkans.

1.2.1.11 Pre-Ottoman and Post-Ottoman Turkic

A number of different Turkic-speaking tribes passed through or settled in the Balkans prior to the arrival of the Ottoman Turks (cf. Girfanova 2001), e.g., Pechenegs, Kumans (Polovtsians), Huns, Avars, and Bulgars. Some of these groups were Kipchak (Northwest Turkic), while others (e.g., Huns and Bulgars), apparently, belonged to a branch whose only living descendant is Chuvash (the so-called *l/r* branch as opposed to the *s/z* branch, to which all other living Turkic languages

belong, e.g., Chuvash *tăxăr* = Turkish *dokuz* 'nine'). The Bulgar Turks (sometimes called *Proto-Bulgars*), who were assimilated by a group of Slavs who took on their ethnonym (the Bulgarians), have left lexical traces in the Balkans (e.g., Sandfeld 1930: 99; Parzymies 1994; Dybo 2012) but, based on the evidence of Old Church Slavonic, no discernible morphosyntactic effects (but cf. Johanson 1996, 2001; Mladenova 2007: 357–362). It has sometimes been claimed that Gagauz, which is generally classed as Oghuz (Southwest) Turkic (as is Turkish), contains Kipchak elements, suggesting possible survivals from pre-Ottoman Turkic peoples (e.g., Kumans, cf. Johanson 1992, 1996, 2002), but these claims are not considered firmly established. It is now generally accepted among historians that the Gagauz are descended from Turkish mercenaries from Anatolia settled in Dobrudja by the Byzantine emperor Michael VIII in the mid-thirteenth century (Wittek 1953; Fine 1987: 215, but cf. also Pokrovskaja 2001). Mollova 1966: 139 notes that Gagauz speakers in Bulgaria consider their language to be a Turkish dialect and observes that speakers refer to their language as *bizim türkçemiz* 'our Turkish.'[31] Finally, we can note here that a significant number of Tatars settled in Dobrudja and many entered Bulgaria in the wake of the Crimean War (1854–1856). Although small groups of Tatars still live in the Balkans today, Tatar has not participated significantly in the long-term processes under consideration here.

1.2.2 Modern Languages of the Balkans

A number of living languages were and/or are spoken in the Balkans but are not considered Balkan languages, i.e., they do not display a significant level of specifically Balkan contact-induced phenomena and generally were not part of the convergence area in terms of linguistic structure. Examples of such languages are presented in the following sections.

1.2.2.1 Hungarian

Spoken throughout the former Hungarian Crown lands, Hungarian is sometimes mentioned in studies dealing with the Balkan languages. For example, Schaller 1975: 151, citing Večerka 1970: 246, connects the Hungarian formation of teens with Slavic influence, e.g., *tisz-en+egy* 'ten-superessive+one = one-on-ten = eleven' (see §6.1.6), but he neglects to mention that the twenties are also formed on that model, e.g., *husz-on+egy* 'twenty-superessive+one = twenty-one,' which could not be a calque on Slavic.[32] We can also note here that Hungarian has

---

31 While it is true that the term *türk* has broad applications in the Turkic group, Mollova is referring here to the perception of Gagauz-speakers, in the context of which *türkçe* refers to local varieties of Turkish.

32 It could, however, be an analogical innovation. Some Aromanian dialects have extended the Slavic pattern to all the tens as well as their teens; thus, *unãsprã-*, extracted from *unãsprãdzace* 'eleven,' is the basis for *unãsprãvingic* 'twenty-one,' *unãsprãtrejdzãts*[i] 'thirty-one' (Gołąb 1984a: 102–104; Koltsidas 1993: 251). Since most of Aromanian preserves the older vigesimal system (*yinghits* 'twenty'), the dialectal Aromanian phenomenon must be a late innovation.

a special treatment of definite direct objects in that it marks them on the verb in its so-called transitive conjugation.[33] Masica 2001: 243–245 links this with Romanian definite direct object marking by means of the preposition *pe*, but he neglects to include the resumptive pronominal clitics (on which see §7.5.1) of Balkan Romance, Balkan Slavic, Albanian, and Greek (and to a lesser extent, Romani), which then connects the region with Turkish and beyond even to South Asia. Hungarian also has definite and indefinite articles, but they are both preposed, as in German (§6.1.2.2.1). In any case, the status of Hungarian with respect to the Balkan sprachbund is basically peripheral, a source of some vocabulary, e.g., Slv *varoš* 'old town,' Rmn *oraş* 'town, town center,' Alb *varosh* 'suburb, neighborhood,' Grk βαρόσι (also βαρώσι) 'suburb, quarter,' Trk *varoş*, all from Hungarian *város* (Sandfeld 1930: 98; Schubert 1999).

1.2.2.2 German

German was spoken in former Habsburg lands and also in colonies of workers brought to Ottoman lands for their technical skills but expelled *en masse* from Yugoslavia after World War Two. Some speakers also emigrated or were expelled from other Eastern European countries. A significant, albeit dwindling, minority continues to live in Romania.

1.2.2.3 Armenian

Armenian was first brought to the Balkans by the Byzantines in the ninth century, with communities coming also from subsequent migrations. The most recent migrations arrived in the wake of the Armenian genocide in Anatolia at the beginning of the twentieth century; there are still Armenian communities in North Macedonia, Bulgaria, and Greece. Their dialects have not received adequate attention, but Adamou 2008 notes that the Armenian of Thessaloniki is considered distinctive in its avoidance of the infinitive.

1.2.2.4 Georgian

A Georgian monastery (P'et'ric'on) was founded at Bačkovo in the Bulgarian Rhodopes in the eleventh century (construction completed 1083), when the Georgian province of Tao and Bulgaria were both part of the Byzantine Empire. Two inscriptions and old copies (thirteenth century) of the monastery's tipikon survive (Šanidze 1971).

---

33 See Aronson 2007 for some discussion that links this Hungarian phenomenon with the Balkan one by way of arguing for an extended contact area.

1.2.2.5 Circassian (Adyghe)

A significant number of Adyghe Circassians were settled in the Balkans by the Ottomans as a result of migrations caused by the Russian conquest of the Caucasus (1769–1864). According to Kurmel 1994 the language was retained in at least one village in Kosovo about 10 km from Prishtina, Čerkes Kjoj (probably the same as current Balshaj (Alb)/Miloševo (Srb)), until the end of the twentieth century, but according to Kumpilova 2012 there were still several Circassian villages in Kosovo as well as some Circassians in Gjilan/Gnjilane up to the 1998/1999 war, when most of those remaining were evacuated to Adyghea. According to Sikimić 2005c, however, there are still some Circassians in Balshaj/Miloševo and in Prishtina. Kânčov 1900: 116, 178, 215 gives locations and statistics for Circassians in Macedonia, but their dialect is completely lost to us (for Kosovo, see also Vermić 2005).[34]

1.2.2.6 Yiddish

Yiddish was spoken in the Austro-Hungarian Balkan lands and by Jewish immigrants from Austria-Hungary, Germany, and the Russian Empire as those states' interests penetrated the southern Balkans. Significant numbers of Yiddish-speaking Jews settled in Wallachia and Moldavia considerably earlier. Ashkenazic Jews were also present in the Balkans from earlier periods. Most were annihilated by the Nazis in World War Two, and most survivors emigrated to Israel, but some communities remain.

1.2.2.7 Other Languages in the Balkans

In addition to these languages, we do not treat the following languages that belong to language groups that are otherwise represented in the Balkan sprachbund.

1.2.2.7.1 Tatar

Although Tatars had expanded from Crimea into Bessarabia and Tatars and Nogais settled in Dobrudja as Ottoman colonists, the majority of Tatars in the Balkans arrived there as a result of the Russian annexation of Bessarabia in 1812 and the Crimean War (1854–1856). The 2011 census in Romania counted 23,935 Tatars. The 1992 census in Bulgaria counted 4,515 Tatars but only 1,803 in 2001. We can also note that Kănčov 1900: 162, 232 records two Turkish villages with the name Tatarli in Macedonia, one in Dojran *kaza* (roughly, 'county'; see footnote 87) and the other in Štip *kaza*.

---

34 The only linguistic datum we have on the Circassian of Kosovo is Bersirov & Tlebzu 1981, who mention the replacement of the ergative by the instrumental. We can speculate that this is a result of Serbian influence.

1.2.2.7.2 Italian

Italian is spoken in parts of Slovenia and Croatia adjacent to Italy and thus outside the Balkan linguistic area as we have defined it; the Venetian dialect of Italian dominated the Adriatic (at times known as the Sea of Venice), and at its height the Venetian Empire included parts of what are now Albania and Greece.[35] Italian is the source of a number of pan-Balkan or widespread Balkan lexical items and remains an important second language in Albania and along the eastern Adriatic Coast.

1.2.2.7.3 Lingua Franca

This was a trade language of the eastern Mediterranean based on Italian but with many other elements (Kahane et al. 1958; Panzac 2008; Nolan 2020; Opferstein 2021). As a language with no native population and one spoken in an area much larger than the Balkans, it is outside the focus of this book except as mentioned in §4.4.3.[36]

1.2.2.7.4 French

Like Lingua Franca, French was used among various Balkan populations as a contact language, and, like Yiddish, it penetrated mostly with Great Power interests. It was not used so much for inter-group contact, however, as for dealing with Westerners, although it was more favored by some groups than by others (see especially Baer 2000; Bunis 1999: 60–122). For political reasons, it remained the language of international contact in the communist Balkans into the 1980s and was particularly important in Romania in the nineteenth century and into the twentieth as a language of scholarship and culture, even providing a model for some grammatical restructuring (e.g., wider use of the infinitive; see Joseph 1983a: 165–166, following Close 1974).

1.2.2.7.5 English

English is currently the most widely spoken contact language for the younger generation in the Balkans. It is a sign of changes in Balkan values that according to the 1994 census in the Republic of North Macedonia (Zavod za statistika na Republika Makedonija 1996), more Macedonians declared a knowledge of English

---

35 Venetian is sometimes treated as a separate Romance language, and in modern times there is an independent Venetian linguistic movement. These issues, however, do not concern us here.

36 We can also mention here a number of so-called secret languages, usually used by craftsmen and tradesmen in Balkan urban centers. These are actually collections of lexical items, or, in some cases, cryptoderivational processes, inserted into the grammar and vocabulary of the host language. While these jargons (in the technical sense) provide interesting evidence for the relative knowledge or lack thereof of the various Balkan languages among different populations, they do not represent independent linguistic systems; see §4.4.3 for discussion.

than of any of the minority languages of the Republic. Such relations are likewise the case in the other modern Balkan nation-states. As Friedman 2019 notes, there is a sense in which, for the Balkans, English is the Turkish of the twenty-first century.

1.2.2.7.6 Slavic Languages

The former Serbo-Croatian is discussed below. Given our linguistic definition of the Balkans, Slovene and various non-South Slavic languages (Rusyn, Slovak, Ukrainian) are all spoken outside our primary area of interest and while some dialects in intimate contact with Romanian may show specific convergent features with the latter (cf. Nomachi 2015 on Slovak), we consider these to be outside our purview here.

1.2.2.7.7 Yet Other Languages

Other languages are spoken by small and isolated populations, e.g., Russian in monasteries on Mt. Athos or among Old Believer groups in Romania. A wide variety of European and Asian languages are spoken by various groups of economic migrants, etc. Such languages are outside our scope.

1.2.3 Modern Balkan Languages – Their Ancestors and Dialects

There is no argument that Albanian, Hellenic, Slavic, Italic, and Indic (or Indo-Aryan) constitute genealogical language groups within the context of Indo-European, nor is there any dispute that Turkic constitutes a genealogically unified group (see Johanson 2021).[37] Within Italic, it is only Latin that went to the Balkans in ancient times, and so all the Balkan languages of Italic origin are in fact Romance. Although the sub-division of Turkic is complicated by overlap of diagnostic features (Poppe 1965: 34–35), the generally accepted distinction of Southwest (Oghuz) and northwest (Kipchak), as well as the *s/z-l/r* division mentioned in §1.2.1.11, is sufficient when discussing the Balkans. The assignment of Romani to a particular sub-group within Indic is an interesting problem (see Turner 1927; Sampson 1927; Matras 2002), but from the point of view of Balkan linguistics it is not crucial. Within the Slavic languages, the separation of South Slavic from North (East and West) Slavic is relatively unproblematic, but divisions within South Slavic require further comment. Albanian and Hellenic have their divisions, and these are treated in §§1.2.3.1 and 1.2.3.2, respectively (see also Joseph 2018 for a concise overview of Greek, both its current state and its historical development).

In any modern account of the Balkan languages, the question of language versus dialect cannot be ignored, and at the same time cannot be definitively solved

---

37 The question of whether the historical relationship of Turkic to Mongolic and Tungusic (and possibly Korean and Japanese) constitutes an Altaic family or an Altaic sprachbund need not concern us. See Johanson 2006 for concise discussion.

without an appeal to heuristic devices and the realities of human identity formation.[38] At the time when the contact phenomena that led to the formation of the Balkan sprachbund as we know it were taking place, religion was at least as important a source of identity as language if not, in some contexts, more so, and glossonymic terminology was neither fixed nor associated with modern nation-states. Thus, for example, during the medieval period, Greek was called Ρωμάϊκα 'Roman,' the referent of *Bolǧar* 'Bulgar' was a Turkic language (see § 1.2.1.11), the language of the various Slavic groups was called simply *Slověnьsky* 'Slavic' (Kantor & White 1976: 74), the term *Vlah* could refer to any form of Balkan Romance (and later also some forms of Balkan Slavic), Roms were called 'Egyptians' or 'Copts' (Turkish *Kıptı)*, Turkish was called *Karamanli* if the speakers were Anatolian Christians, and the glossonym of the Albanians underwent a transformation from *Alban-/Alvan/Arvan-/Arbër-* to *Shqip* (Hamp 1994a: 66).[39] The extension of ethnonyms like *Serb* and *Bulgar* to refer to languages did not take place until later in the Middle Ages, well after the Slavic migrations, while Modern Greek did not take on the label "Hellenic" until West European Romantic ideas penetrated the Balkans in the late eighteenth and nineteenth centuries (cf. Herzfeld 1982, 1987).

There are instances when the scientific knowledge of the linguist can and must be opposed to politically motivated ideologies, e.g., the claim that Modern Macedonian is not Slavic, Modern Slovene is descended from Venetic (which

---

38 The distinction between so-called *abstand* ('distance') and *ausbau* ('development') languages can be noted here. The terms as introduced by Heinz Kloss supposedly distinguished standard languages that were inherently independent, e.g., Greek, from those where two or more dialects on a continuum were standardized, e.g., Bulgarian and Macedonian. See Tosco 2008 for an overview of research and Fishman 2008 for a salient critique of Kloss's distinction. We can note in passing, as does Fishman, that any standard language involves the intervention of human agency and that diachronic processes of dialect differentiation and language change will affect the development of any standard. Bugarski 2002 introduced the term *umbau* (literally 'renovation, rebuilding') to describe the break-up of Serbo-Croatian, which, for the most part, did not involve the elaboration of distinct dialectal bases. Another aspect of the problem of the definition of language is the question of number of speakers. Various sources claim higher or lower figures or give no statistics at all, depending on perceived national or other interests, groups are underreported or over-reported owing to social or political pressures, figures for speakers in diaspora may or may not be included, and so on. Since our concern here is with linguistic processes as such, we do not attempt to estimate total numbers of speakers. See Friedman 1996b and 2003b for a discussion of such issues in connection with Macedonian censuses.

39 At the time of the Slavic invasions of the Balkan peninsula (c.550–c.630 CE, cf. Fine 1983: 25–73), the various Slavic-speaking tribes did not have the type of modern national identities sometimes projected back onto them in modern works (e.g., Franjolić 1983; cf. also Banac 1984: 189; see Fine 1983: 33–37, 49–59 for an objective account). The evidence that we have indicates that during the early Middle Ages, despite tribal and territorial divisions, they thought of their language as an entity we can call *Slavic* (or *Slavonic*). Thus, for example, in the *Vita* of Methodius, referring to events in the ninth century (although our manuscript is three or four centuries later), Rostislav and Světoplŭk I of Moravia refer to themselves and their people as "*my Slověne*" 'we Slavs' and the Byzantine Emperor Michael III says to Constantine the Philosopher: " . . . *Solounane vьsi čisto slověnьsky besědouioṭъ*" ' . . . the Thessalonians all speak pure Slavic' (Kantor & White 1976: 74; see also Darden 2004). Indeed, had the Slavs not spoken essentially the same language during this period (although, to be sure, we know that some dialectal differentiation had already taken place), the mission of Constantine (Cyril) and Methodius could hardly have succeeded nor would their language have been accepted throughout the Slavic Balkans (cf. Fine 1983: 49–59).

is definitely non-Slavic and accepted by Hamp 1994b as Italic, although Wallace 2004: 842 still considers the question of Venetic's precise position within Indo-European to be open), that any human language can exist unchanged for thousands of years, or that any living language (or its immediately reconstructible ancestor) is the source of all human language. Similarly, accounts of the formation of the Balkan standard languages, processes which began precisely at the time when linguistics was emerging as a modern scholarly discipline (Friedman 1997a) and when the earliest phases of what would become Balkan linguistics also began (see Chapter 2), must take into consideration the problems of essentialization and reification inherent in any attempt at valorizing some processes as somehow "natural" and others as "artificial" (cf. Irvine & Gal 2000; Gal & Irvine 1995; and Gal & Irvine 2019, as opposed to Kofos' 1993 perverse interpretation of Anderson's 1983 *imagined* as *imaginary* when applied to Modern Macedonian versus Modern Greek linguistic identity formation).[40] To this can be added the fact that mechanistic attempts at counting "Balkanisms" (see §3.3 and §3.4.2.2) and assigning values to languages or linguistic features will be skewed by the counter's definition of *language* and of *Balkan*. As stated at the beginning of this chapter, the fact that boundaries occur in nature can affect or be ignored by linguistic processes. Similarly, while the establishment of an isogloss – e.g., the boundary between the region where Common Slavic back nasal /ɔ̃/ gives modern /u/ as opposed to a vowel that is not high and/or rounded, or the boundary between the region where a Common Slavic demonstrative pronoun developed into a postposed definite article as opposed to the region where it did not – can be treated as a relatively straightforward mapping process, whereas the assignment of linguistic significance to such isoglosses (aside from determining their relative chronology) becomes basically arbitrary, especially in the context of geopolitical territorial claims.[41] In the case of these two examples, the first isogloss is associated with Serbian and extends eastward to about longitude 23 and south roughly to latitude 42, while the latter is associated with Bulgarian and extends westward to about longitude 22 and north to around latitude 43.

1.2.3.1 Albanic

There is no doubt that Albanian is descended from one of the poorly attested or unattested ancient languages of the Balkans, but, as is clear from the discussion in §1.2.1, the exact identification cannot be made with any certainty (but see §1.2.1.2, footnote 18). Hamp 1994a argues on the basis of shared innovations such as the Winter's Law vowel lengthening (see footnote 18) that within Indo-European, Albanic was closest to Balto-Slavic (see Hyllested & Joseph 2022 on Albanic

---

40 The natural/artificial distinction was used extensively in the 1990s to justify the success or failure of specific manifestations of the same human political construct: the nation-state.

41 Although even here there can be complicating details, cf. Alexander 2000a, 2013 or any number of dialect atlases that give distributions of individual lexical items containing reflexes of what was, at an earlier period, the same sound in the same environment.

forming a subgroup with Greek) and this shared innovation together with shared ancient non-Indo-European borrowings such as the word that became 'apple' in Germanic and Balto-Slavic but 'sorbus' in Albanian – *vadhë* (Geg *vodhë, vollëz,* etc.) – indicates that during the late Indo-European period, the dialect that ultimately became Albanian was part of the North (Central) European area (Hamp 2010). At some later point, Albanoid, i.e., the Indo-European language that became Albanian prior to its contact period with Latin, was spread along the Carpathians from southeastern Poland to Croatia, as attested by such widespread lexical items as *vatra* 'hearth' (Hamp 1976, 1981), *strungă* (Rmn)/*shtrungë* (Alb) 'enclosure for milking animals' (Hamp 1977b). Hamp 1999 uses the term Proto-Albanian for the period of contact with Latin and describes this period in the following terms: "The eastern portion of this speech area adopted Latin but kept traces of its old grammar and many lexemes. The result is called Romanian. [. . .] The western portion accepted many loans but kept its language" and became Albanian (Hamp 1994a: 67). In Hamp's terminology, Common Albanian refers to the period after the contact with Slavic but before the dialectal split between Geg and Tosk. Attested Albanian begins with a baptismal formula embedded in a Latin text from 1462, and the first major dated textual attestation is the 1555 *Meshari* 'Missal' of Gjon Buzuku (Çabej 1968).[42]

The Geg/Tosk split is defined by a relatively compact bundle of isoglosses whose northern border (rhotacism and the development of *uo into *ue* (Geg) or *ua* (Tosk) in syllables closed by a sonorant) runs along the course of the river Shkumbî in central Albania and whose southern edge (denasalization and the development of stressed schwa) is roughly 10–20 kms to the south (with the change of **vo-* > *va-* located in between).[43] It represents a defining moment in the history of Albanian unlike any other since contact with Latin. Although there are other features participating in the split (e.g., infinitive with *me*, sigmatic imperfects), the key phonological features are both emblematic and, in a certain measure, identifiable in temporal terms. That the split post-dates contact with Latin is certain, since Latin loanwords uniformly undergo the changes that characterize the split, e.g., *orphanus* > *vorfën/varfër* 'poor,' *(h)arēnus* > *rânë/ rërë* 'sand.' Equally certain is that the changes had occurred before the diasporas

---

42 There have been various claims and attempts to discover or produce Albanian texts from earlier periods; see Elsie 1986, 1996 for discussion. We can also note that in 2002 an old Albanian manuscript was claimed to have been discovered in the Vatican library, but to date the actual text has not been authenticated or even produced. There is also an undated pericope of Matthew 27: 62–66 in Tosk written in Greek letters and referred to as the Easter Gospel. It is of unknown date or provenance. The oldest documents date from the Early Modern period (sixteenth to eighteenth centuries) and the language of these documents is referred to in this book as Early Modern Albanian, rather than Old Albanian, as is done in other works. Should we ever be so fortunate as to find manuscripts from the medieval or earlier periods, then the adjective "old" is appropriate, but here we agree with historians, who recognize the relevant centuries as "early modern."

43 It is interesting to note that in earlier times, this transitional zone narrowed to a single bundle at the Black Drin River and followed the course of the river through the middle of the town of Struga on the north shore of Lake Ohrid in today's Republic of North Macedonia. Albanians living on the left bank of the river spoke Geg, while those on the right spoke Tosk. Population movements have eliminated this distinction, but it was still in effect during the earlier part of the twentieth century.

of the late medieval and early modern periods that produced Arvanitika and Arbëresh, which are unmistakably Tosk (see below). Between the *terminus a quo* and the *terminus ad quem*, however, stretches a period of seven or eight centuries during which another event significant for the history of Albanian occurred: the arrival of the Slavs.

There are three main proposals concerning the chronology of the Geg/Tosk split: (1) It was completed before contact with Slavic (Gjinari 1989: 2); (2) It began prior to but continued into the period of contact with Slavic (Hamp 1994a; Rusakov 2013); and (3) It began after contact with Slavic (Janson 1986). Without entering here into the details of the arguments adduced, we note simply that the evidence of Slavic loanwords for the chronology of the Geg/Tosk split is sufficiently meager and complex that any given account must explain away counterexamples. On the one hand, Slavic loans in /n/ that do not rhotacize in Tosk can be explained as analogical or late (e.g., suffixal *-nik*, *stopan* 'herdsman') while on the other, rhotacized examples of obvious antiquity such as Geg *shtëpâ/shpneshë* 'herdsman/housewife,' the latter equivalent to Tosk *shtëpreshë* 'housewife,' from Slavic *stopan* 'master, householder,' can only be explained away by claiming a non-Slavic intermediary (e.g., the ancestor of Romanian *stăpîn* 'master, owner') or by arguing that rhotacism continued as a local phenomenon in isolated areas. If, with Hamp 1994a: 67 and Jokl 1923: 42, 46, we accept the treatment of Slavic *č as *s in Albanian *porosit* 'order' from Slavic *porǫčit-* as evidence that initial contact with Slavic took place prior to or during the Common Albanian period, then the issue of the exact dating of rhotacism, however important it may be for Albanian etymology, becomes moot for the question of Slavic-Albanian language contact, since by definition Common Albanian pre-dates the Geg/Tosk split (see also Rusakov 2013 on the relative dating of Albanian isogloss bundles.)

Although other, more recent, isoglosses sometimes traverse the Shkumbî (Maynard 2002; Friedman 2003f), Albanian dialectological divisions are basically subordinate to the Geg/Tosk distinction. Within Geg we can differentiate Southern Geg, East Central and West Central Geg, Northeast Geg, and Northwest Geg. Southern Geg stretches from the Adriatic to just beyond Mts. Jablanica and Belica in North Macedonia and includes Elbasan and Tirana. This dialect functioned as the basis of a *de jure*, but not always *de facto*, Albanian standard from 1923 until the early 1950s. West Central Geg is entirely in Albania, north of Southern Geg, and East Central Geg begins in Albania just west of Mts. Korab and Jablanica and then extends into North Macedonia, accounting for the great majority of Albanian dialects there from Tetovo and Kumanovo almost to Bitola.[44]

---

44 According to Rexhep Ismajli (p.c.), the old Ohrid Albanian families are Geg-speaking, whereas the Gjupci/[Balkan] Egyptians of North Macedonia speak Tosk. This last point has been verified in Friedman's own fieldwork. During Friedman's fieldwork in North Macedonia in 1972, these latter called themselves Gjupci. By 1981, however, they were calling themselves Egipkjani (Friedman 1985c). They are among the groups of Romani descent who do not speak Romani and who do not identify as Roms (cf. some Romanian-speaking Banjaš in Serbia, the Turkish-speaking Millet of Bulgaria, the Albanian-speaking Ashkali of Kosovo, etc., Sikimić 2005a; Marushiakova et al. 2001; Krumova et al. 2004). Note that in North Macedonia, Muslim Balkan Egyptians speak Tosk

The dialects of Montenegro and Shkodër are Northwest Geg, while Kosovo and adjacent parts of Albania and some Macedonian border villages speak Northeast Geg. The only older Geg diaspora dialect is that spoken in the village of Arbanasi (Itl Borgo Erizzo), a suburb of Zadar, to which inhabitants of two villages near Bar (Itl Tivari) emigrated in the eighteenth century. There is also evidence of a Geg dialect in Istria (Altimari 2011a) and Srem (Hamp 1994a). The Northern Tosk dialects occupy most of southern Albania and extend into the southwestern corner of the Republic of North Macedonia and much of the western part of northern Greece. They also serve as the basis for modern Standard Albanian (Byron 1976). The Lab dialects are spoken south and west of the Vjosa River and extend into Greece between Northern Tosk and Çam. The Çam dialects begin between Butrint and Konispol in southern Albania and extend into Epirus/Çamëri. Although exempted from the exchange of populations between Greece and Turkey as arranged in the Treaty of Lausanne in 1923 (despite the fact that they were Muslims),[45] most Çams were expelled from or left Greece in the wake of the Civil War in 1948, although there are still remnants of the Albanian-speaking population in Epirus (Rexhep Ismajli, p.c.).

There are three main Tosk diaspora dialects. Arvanitika separated from southern Çam by the thirteenth century or earlier and was spoken all over southern Greece and on many of the islands into the twentieth century (see Hamp 1961). The language, which its speakers call Arbërisht, has undergone massive attrition under pressure from Greek since World War Two and has disappeared completely in places where in the 1950s there were still significant numbers of speakers (Hamp, p.c., see also Tsitsipis 1998). Arbëresh, still spoken in forty-five to fifty villages of southern Italy (Altimari 1993; Hamp 1994a; Nasse 1964), separated from Arvanitika and Albanian under pressure from the Turkish conquest, especially after the death of Skanderbeg 1468 and after the reprisals for the rebellion of 1481. Another exodus to Italy occurred in 1492 (Fine 1987: 602). The one other set of migrations (which probably involved more than one wave) was eastward mostly from the northern Tosk region, but with some southern Tosk elements (Hamp 1965; Liosis 2021) also in the early modern period, to what is today Bulgarian, Greek, and Turkish Thrace and to northeastern Bulgaria, and some speakers continued from northeastern Bulgaria to Southern Dobrudja (Budžak) and Crimea in the nineteenth century. Some Albanian speakers, who found themselves on the Turkish side of the border after the partition of Thrace, migrated across the border to the Greek side (Friedman 2004a: 59–155; Novik et al. 2016: 39–358; Dalatsis 2016; Kotova 2017; Johalas 2019; Liosis 2021). These speakers, like the Albanian speakers of Albania and contiguous areas, generally call their

Albanian and Christians speak Macedonian while in Kosovo Balkan Egyptians and Ashkali speak Geg. Cf. also the difference between usages of *Tsinganos* and *Yiftos* as autonyms among Roms in Greece; see Messing 1981; Hunt 1999.

45 Of the Muslims in Greece, only the Çams and the Muslims of Thrace (mainly Turks, Roms, and Pomaks) were allowed to remain. In exchange, the Greek Orthodox of Constantinople were also allowed to remain in Turkey. On the exchanges of populations, see Ladas 1932; Pentzopoulos 1962; Hirschon 2003; Clark 2006; Yıldırım 2006.

language *shqip*, which is probably related to *shqiptoj* 'pronounce clearly', cf. Latin *explicare* (Hamp 1994a). However, some villages simply use *si neve* 'like us' (cf. the Macedonian speakers of Boboshtica and Drenova (Boboščica and Drenov^jäne, in the local dialect), in Albania who use *kaj nas* 'like us'). The villages of Léhovon (Mac Lehovo), Drosopigi (Mac Bel Kamen), Flámbouro (Mac Negovani), and Eláteia (Mac Elovo), south and east of Flórina (Mac Lerin) in what is now Greek Macedonia, were settled by Christian Albanian- and Aromanian-speakers from Epirus at the beginning of the nineteenth century (Simovski 1998[2]: 137–182). At one time there were also Albanian villages across the Rhodope range and into Thrace, then up the Black Sea Coast to the Danube. Of these, a single village in Bulgaria, Mandrica, near the Turkish and Greek borders, as well as some neighboring villages on the Greek and Turkish sides, survive (Hamp 1972b; Sokolova 1983; Stankov 2016; Liosis 2021). An offshoot of Mandrica from the Balkan Wars is Mandres, south of Kilkis (Kukuš) in Greek Macedonia (Hamp 1965). Three villages of Albanians were also settled near the Sea of Azov in Ukraine in 1862, after the Crimean War (Kotova 1956: 254–255). See Friedman 1994c for a discussion of the dialects of Turkish Thrace and Liosis 2021 for Greek Thrace. Many Albanian-speakers left what is now former Yugoslavia for what is now Turkey as the Ottoman Empire contracted, especially in 1878–1923, and many more left Yugoslavia in the 1950s. Their dialects survive to varying degrees.[46] See also Desnickaja 1968 on Albanian dialects.

1.2.3.2 Hellenic

The Greeks came to the Balkans during the first half of the second millennium BCE, displacing or absorbing other Indo-European and/or non-Indo-European peoples (Browning 1983: 1; Drews 1988; Gindin 1967; Horrocks 2010: 9).[47] Although Greek was first written using the Mycenaean syllabary (so-called Linear B) in the second half of the second millennium BCE, the syllabary was lost with the collapse of Mycenaean civilization (c.1200 BCE) and the Greeks reverted to illiteracy. Eventually the Phoenician alphabet was adapted to the representation of Greek, and by the eighth century BCE inscriptions in Greek are to be found and there are as well works that can be called literary (Browning 1983: 3). From then on into the Hellenistic and Roman periods (c.300 BCE to 300 CE), there was a flowering of production of literary, historical, and scientific works in Greek, and literally thousands of inscriptions of all sorts (official decrees, treaties, dedicatory and funerary epigraphs, informal graffiti, and so on) were

---

46 See Andrews 1989: 130–133 for current data on Albanians in Turkey, Ellis 2003 for discussion of the migrations after World War Two, and Geniş & Maynard 2009 on the establishment of Albanians in Samsun Province after the Balkan Wars of 1912–1913. There are also many important Albanian émigré communities in Western Europe, the United States, and Australia.

47 There is some controversy as to the extent of uniformity in the language at the point of entry, whether the familiar Ancient Greek dialects formed once Greeks were in the southern Balkans, and so on, but these issues are largely irrelevant to our concerns here. See, e.g., Buck 1955 for the traditional view of the Ancient Greek dialects.

produced. By the first century CE, however, as the result of influence from (and reverence for) classical models of Greek usage, written Greek had become sufficiently divorced from the spoken language that for more than a millennium much of the textual evidence is never better than equivocal in terms of representing contemporary usage. Nonliterary papyri, however, containing personal letters, among other types of documents, do provide some insights into the language as found in more mundane day-to-day uses up to approximately the seventh century CE. In the twelfth century we begin to find texts such as the poems of Ptochoprodromos and Michael Glykas that do not attempt to imitate in a puristic fashion earlier literary (generally Attic Greek) models and can be taken as more reliable representations of what the spoken language may have been like. This is generally viewed as the beginning of Modern Greek. Nonetheless, even these late medieval and early modern texts have features whose status as puristic versus vernacular cannot be established with certainty. The Ερωτόκριτος of Vintsentzos Kornaros of seventeenth-century Crete represents the culmination of this tradition (Browning 1983: 7–9). The competition between the consciously archaizing and Atticizing Katharevousa (Puristic) Greek and the vernacular-based Dimotiki (Demotic) Greek, begun in the nineteenth century, has been decided in favor of Dimotiki, though the influence of Katharevousa on Dimotiki is undeniable, mainly in matters of vocabulary, phonology, and morphology. Nonetheless, Katharevousa is the language of an important body of literature and continues to interact with Dimotiki (on which, see Mackridge 1985, 1990; Kazazis 1992, among others).

Although Ancient Greek was divided into a number of dialects, the Koine ('common') dialect, based on Athenian Attic with significant input as well from the Ionic dialect, had displaced all of them by the end of the second century CE (Browning 1983: 50–52). The only substantial remnant of any other Ancient Greek dialect is Tsakonian, which is descended more or less directly from Ancient Doric (the Laconian variety) and spoken in the mountains of the southeastern coast of the Peloponnese and, until the exchange of populations between Greece and Turkey, on the southern shores of the Sea of Marmara by colonists from the eastern Peloponnese who had migrated there in the fifteenth century (Browning 1983: 124).[48] There are also some ancient Doric elements in the Greek of southern Italy . All remaining dialects of Greek are descended from the Koine. On mainland Greece, another wave of dialect extinction began with the end of Turkish rule in the nineteenth century. Since the Peloponnese constituted the overwhelming majority of the territory of the early Greek state (and most of Attica and Central Greece was heavily Arvanitika-speaking), immigrants from the Peloponnese poured into Athens, the new capital, and the town dialect of Athens together with the dialects of Megara, Aegina, and Kyme in Euboea (the so-called Old Athenian dialect) which,

---

48 According to Scutt 1912–1913: 142, the syntax and structure of Tsakonian are very similar to Modern Greek, although the present and imperfect have innovative analytic formations using 'be' plus a participle (Scutt 1912–1913: 168). The most authoritative description of Tsakonian to date is Pernot 1934, but see now Kisilier 2021 on earlier contacts of Tsakonian with Greek; in more recent years, Tsakonian has been influenced considerably by standard Modern Greek.

among other features, had /u/ from Ancient Greek /ü/ (upsilon), was submerged in a flood of Peloponnesian.

The Peloponnesian dialects are relatively homogeneous, and the dialects of Istanbul and the Ionian Islands are relatively close to Peloponnesian (Mackridge 1985: 5). Of the central or core dialects, this leaves only the dialects of northern Greece as distinct.[49] According to Browning 1983: 120–121, the main isogloss separating the northern dialects goes along the coast of Epirus and Acarnania, along the Gulf of Corinth and across the Isthmus, along the northern mountain frontier of Attica, south of Euboea, across the middle of the island of Andros, north of Icaria and south of Samos (but excluding Chios) and to the coast of Asia Minor. North of this line, unaccented high vowels (/i/ and /u/) are dropped (sometimes with palatalization of a consonant before unaccented /i/) and unaccented mid-vowels (/e/ and /o/) raise to high vowels. Northern dialects also have velar [l] before back vowels, pronounce /s/ as [š] before front vowels, and use the accusative, in some instances preceded by από 'from,' in place of various uses of the genitive.[50] It is worth noting that the north is where settled populations speaking Albanian, Aromanian, Macedonian, Bulgarian, and Turkish (as well as Romani, Judezmo, and Meglenoromanian) were often in the majority, or plurality, depending on the region or village, and these distinctive features of northern Greek, if not contact-induced, are at the very least consistent with the larger linguistic environment, e.g., some of these same phonological processes occur in Albanian, Aromanian, and Balkan Slavic (in this region).

The exchange of populations with Turkey in 1923 resulted in the expulsion from Greece of 500,000 Muslims speaking all the languages of the Balkans, mostly from northern Greece, and the forced resettlement of 1,500,000 Orthodox Christians, many of them Greek-speaking but also speakers of Turkish, Romani, etc. The Christians from towns like Smyrna (Trk Izmir) were settled in the outskirts of Athens, while the Christians from rural areas such as Cappadocia and the Pontos (Black Sea) region, who were more likely to speak only Turkish, were settled in Aegean (Greek) Macedonia and other parts of northern Greece (see Ladas 1932; Pentzopoulos 1962). The new immigrants and their children assimilated to standard Greek although the older generations in Greek Macedonia also learned Macedonian.[51]

The dialects of Mainland and Aegean Greece form a center in relation to which the remaining dialects constitute a periphery of varying degrees of antiquity and differentiation. Starting from the west, there are two enclaves in southern Italy, one each in Apulia and Calabria, which continue the dialect of Magna Graecia of the

49 It is worth noting that the Greek dialects of Albania do not display uniformly the features of northern Greek dialects. One interpretation is that those isoglosses did not extend beyond the border with Albania, but it could also suggest that the Greek of southern Albania was spread entirely via education and the church during the course of the past two or three centuries among speakers of other languages in that region (see C. Brown & Joseph 2013 on these possibilities).

50 Another important isogloss pertains to nasal + voiced stop combinations; see §5.4.4.1 for discussion.

51 Miladina Monova (p.c., also VAF field notes).

late Roman Empire (Browning 1983: 132), Crete, the Cyclades, the Dodecanese, Cyprus, Asia Minor, Ukraine (Rostov region and Mariupol in the Azov region, whose speakers came from Crimea), and the Caucasus (Abkhazia and southern Georgia). The Asia Minor dialects consisted of Bithynian in the northwest and Pontic in the northeast, related to dialects near Rostov in Ukraine, Livíssi (Trk Kayaköy) in the southwest, and Cappadocia, Phárasa (Trk Çamlıca), and Silli (Trk Sille) in central Anatolia.[52] Most speakers of Anatolian Greek dialects were resettled in Greece during the exchange of populations, and except for a few Pontic-speaking Muslim villages in the northeast, these dialects have disappeared from Turkey.[53] Cappadocian, however, once thought to have died, is still alive and in use in villages in northern Greece (Janse 2009). Most speakers of Greek in the Caucasus left for Greece after the collapse of the Soviet Union and subsequent war in Abkhazia. The dialects of Mariupol and Anatolia show, as might be expected, influence from Tatar and Turkish, respectively. Thus, for example, Cappadocian Greek has the same eight-member vowel system as Turkish, while Mariupol Greek has lost the genitive and expresses possession by means of an *izafet* construction, e.g., σπίτι-τ πόρτα 'door of house,' τάτα-του σπιτ 'his father's house' (Browning 1983: 135–136). As with the diaspora dialects of other languages, for the most part we do not consider these dialects here.[54]

Two Greek ethnolects that should be mentioned are those of the Sarakatsans, transhumant shepherds (now mostly settled) living in northern Greece, Bulgaria (where they are known as Karakačani), and the Republic of North Macedonia (where they are known as Sarakačani) and the Romaniote Jews, whose ancestors were living in Greece prior to the arrival of the Sephardim of the Iberian peninsula (see §1.2.3.3). Owing to their lifestyle and material culture, there had been speculation that the Sarakatsans were Hellenized Vlahs (see §1.2.3.3). As Høeg 1925–1926 demonstrated, however, the dialect of the Sarakatsans of Northern Greece is a straightforward northern Greek dialect that gives no indication of any unusual interference from Aromanian. Tzitzilis 1999 confirms that the dialect of the Karakačans of Bulgaria is of the same type. Most of the Sarakačans of what is now North Macedonia left for Greece during the 1960s owing to attempts at collectivizing their flocks (Nedelkov 2011).[55]

52 See Dawkins 1916 on Greek of Asia Minor, especially Cappadocian, and more recently Janse 2019; see Drettas 1997 on Pontic, and Andriotis 1961 on the Greek of Livíssi (Trk Kayaköy). See also Joseph 2003a on the Pontic Greek of the Black Sea area in the former Soviet Union and Kisilier 2009 on the Greek of the Azov region.
53 See Mackridge 1987 and Sitaridou 2013, 2014a, 2014b on these dialects.
54 Modern Cypriot Greek, though part of the Southeastern dialect group, is perhaps a special case as it had its own medieval literary tradition. Note, e.g., the fifteenth-century *Chronicle of Makhairas* (Dawkins 1932) and the fifteenth-century *Chronicle of Boustronios* (Kehiagioglou 1981 and cf. the translation of Dawkins 1964). The local variety is quite different from mainland Greek in aspects of its phonology and weak object pronoun placement but nonetheless shows the same diagnostic morphosyntactic Balkan features as mainland dialects, for which sustained contact over the years with Balkan Greek may well be responsible.
55 See also Kahl 2007.

The Greek dialect of the Romaniote Jews is known as Yavanic (Yevanic) or Judeo-Greek. For the most part, Yavanic was supplanted by Judezmo after the arrival of a large number of Sephardim from Spain in 1492 and Portugal in 1496–1497, but some communities preserved the Romaniote liturgy, and Greek-speaking Jews produced written texts in Constantinople in the sixteenth century (see Hesseling 1897 for an edition of a 1547 translation of the Pentateuch into Greek, prepared for didactic purposes to help Jewish Greeks learn Biblical Hebrew).[56] By the beginning of the twentieth century, Jews in the towns in Epirus of Ioánnina, Árta, and Préveza, and in Chalkída still spoke a form of Greek that differed in some features from the Greek of their Christian neighbors. The differences seem to be limited to phonetic, intonational, and lexical phenomena. In contrast to some other Jewish languages, no awareness of language separateness seems to have existed. As with many other Jewish communities with distinctive languages or ethnolects in Europe, the Holocaust destroyed the majority, most of the survivors went to Israel, and the language is in a state of attrition. See Connerty 2003 and Krivoruchko 2011 for studies of Yavanic.

One problem with the study of Modern Greek dialects is the relative paucity of material, especially in the more generally accessible languages. Moreover, for the most part, the most important primary descriptive and analytic sources treat island or diaspora dialects and thus are less useful with regard to Balkan Greek. In fact, as Kontosopoulos 1981: 131–132 points out, the dialects of much of mainland Greece, including Macedonia and Thrace, still have not been adequately described. Some notable exceptions are Mirambel 1929 and Pernot 1934, though they deal with the Peloponnesos, and useful summative works include Thumb 1912, Dawkins 1940, and most importantly Newton 1972 (with extensive reference to the relevant literature in Greek); Trudgill 2003 is an important reassessment of the dialect divisions of Modern Greek, and Tzitzilis 2022 offers a useful handbook survey (in Greek) of each of the major dialects.

1.2.3.3 Balkan Romance[57]

Prior to Augustus (27 BCE–14 CE), Roman influence in the Balkans was limited to the coast, particularly the coastal towns. From the second century BCE through the first century CE, the Romans gradually annexed most of what is now the Balkans south of the Danube, and the second century CE saw their relatively brief (107–271) occupation of Dacia (roughly, modern Transylvania and western Wallachia). Studies of the language of inscriptions indicate that Roman linguistic influence extended southward to the so-called Jireček line (Jireček 1911). Petar Skok identified a somewhat different boundary, and the space between the two is now presumed to have been a zone of Latin-Greek bilingualism (Rosetti 1964a: 34–36).[58] According

---

56 Krivoruchko 2008, 2014 offer more recent appraisals of this linguistically important work, and Joseph 2019a adds further discussion of the Jewish Greek of Constantinople based on this text. See also §3.2.2.10 and §7.7.2.1.1.2.1, footnote 144 for more on this text and what it can tell us.

57 For a variety of papers on Balkan Romance, see Sikimić & Ašić 2008.

58 On modern approaches to ancient bilingualism, see Adams et al. 2002; Adams 2003; Rochette 2010, and references therein.

to Jireček, the line began at Lezha (ancient Lissus) while Skok identified the southern limit of Latinity as beginning at Vlora (near ancient Apollonia). Jireček's line then went east across Albania and North Macedonia, between Skopje (Scupi) and Stobi (near Prilep) south of Niš (Naissus) and Pirot (Turres) but north of Sofia (Serdica) and then across the Balkan (Haemus) range to Varna (Odessos). Skok's line went northeast south of Ohrid (Lychnidus) and Skopje (Scupi) and north of Sofia (Serdica) to cross the Balkan range to Varna (Odessos). The Black Sea coast, however, was dominated by Greek as far north as Tulcea (Aegyssus) in today's Romania (cf. Rosetti 1964a: 34–36, 1973: 47–48; Kaimio 1979: 86–89). South of these lines, the dominant language was Greek. East Balkan Romance in its various forms (Romanian, Aromanian, Meglenoromanian, and Istro-Romanian) is descended from the language of Roman colonists and Romanized peoples east of Dalmatia and, in all likelihood, both north and south of the Jireček line, since the Romans maintained garrisons to guard roads throughout the Empire. A contested issue between the Romanians and Hungarians, however, is the question of whether the Romanian of present-day Romania, especially Transylvania, is descended from the language of Romanized Dacians and colonists who remained behind after Trajan's evacuation of 271 CE (and is thus 'autochthonous'), or whether it is descended from the language of Romans and Romanized peoples living south of the Danube who did not cross over the Danube (or at least into Transylvania) until after the arrival of the Magyars in the late ninth century (cf. Fine 1983: 10; Saramandu 2003–2004 argues for Romance continuity in Romania, while Du Nay 1996 argues against it).[59] Contributing to this problem is the enormous gap between the last Roman inscription (late sixth century CE, Minkova 2000) and the first dated surviving document in Romanian, the Letter of Neacşu of Câmpulung from 1521.[60] As a result, the arguments on both sides are based on conjecture and circumstantial evidence of uncertain quality.

Among Romanian linguists, there is a disagreement between those who recognize Aromanian as a separate Balkan Romance language and those who would make of it (along with all of Balkan Romance) a dialect of Romanian despite the many differences and the fact that the two have been separated for about a thousand years.[61] Aromanian (also known as Vlah, see below) is spoken south of the Danube in modern-day southern Albania, northern Greece, North Macedonia, and adjacent parts of Bulgaria as well as by émigré colonies in Dobrudja and elsewhere. See Ivănescu 1980: 30–46 for a summary of the debate; cf. also Savić 1987; Bacou 1989; Peyfuss 1994; and Jašar-Nasteva 1997. Aromanian is recognized and used as a distinct language in the Republic of North Macedonia.

59 Saramandu is a Romanian linguist of Aromanian origin. Du Nay is the pseudonym of Adam Makkai, an American linguist of Hungarian origin.

60 Recent work using watermarks has shown that *The Hurmuzaki Psalter* was copied around the year 1500 (Mareş 2000, cited in Timotin et al. 2016: 1). Note that in this book we use the term Early Modern Romanian rather than Old Romanian. The arguments are the same as those for Early Modern Albanian cited in footnote 41. Cf. also Nicolae 2015.

61 Close 1974: 67 notes that there were early attempts to create a literary Romanian which combined Romanian and Aromanian dialects, but these were abandoned as impractical.

We should also note here the existence of Meglenoromanian, surviving in a handful of villages in the southeast of the Republic of North Macedonia and adjacent parts of Greece as well as among migrants in some of the towns.[62] At the beginning of the twentieth century, Meglenoromanian was spoken in about a dozen villages in what was then the Ottoman *kaza* (county) of Gevgelija, *nahiye* (township) of Karadžova, on territory that was divided between Greece and Serbia (eventually the Republic of North Macedonia) in 1913 (Atanasov 1990: 1–14; one village in the region, Livãdzi, was Aromanian-speaking, its inhabitants having arrived in the eighteenth century (Puşcariu 1976: 224)). The largest Meglenoromanian village, Nănti (Mac Noti, Grk Nótia) was Muslim and ended up within Greece's borders. With the exception of a single family that converted to Christianity, of whom only a single member was still alive in 1984 (Atanasov 1984: 479), almost the entire village was sent to Turkey during the exchange of populations in the 1920s and settled in various parts of eastern Thrace and western Anatolia (see Kahl 2006).

Linguistically, Meglenoromanian is heavily Slavicized (e.g., it has borrowed the prefixal system of Slavic aktionsart), showing evidence of long and intense contact with Macedonian. This is in contrast to Aromanian, which, in its various dialects, shows significant influence from contact with Greek and Albanian as well as Macedonian. The most important historical linguistic question raised by Meglenoromanian, however, is whether it represents the language of a population that became linguistically separated from Common Balkan Romance at the same time as Aromanian or at a later date. Although Atanasov 1999 argues that Meglenoromanian represents a later break-off from Common Balkan Romance that arrived via the Morava and Vardar valleys rather than via the Rhodopes (this latter route was posited by Capidan 1943: 16–17), his chief arguments rest on shared archaisms (e.g., a preserved infinitive) rather than shared innovations, and such shared innovations as are cited could represent later parallel developments.[63] In general, shared innovations link Meglenoromanian to Aromanian, e.g., the change of velars to dentals before front vowels (Todoran 1977: 102–109), although the separation probably occurred at an early date (see also Kahl 2006).

The term *Vlah* entered the Balkans via a Gothic (and therefore Germanic) intermediary which had it from a Celtic tribal name (Skok 1973: 606–609). It is recorded by Caesar as *Volcae*, by Strabo and Ptolemy as *Ouólkai*, and it was in the transfer to Gothic (as **walhs)* via Latin that the ethnonym took on the meaning 'foreigner' or 'those folks over there' or 'Romance speaker' (and, later, also 'transhumant shepherd' and other meanings). The metathesis of *Wal-* to *Vla-* is typically South Slavic. In Greece, the use of Βλάχος to mean 'shepherd' is

---

62 See Friedman 2009c for a detailed list of toponyms and a map.

63 An example is the change of /d/ plus front vowel to a voiced dental affricate in Aromanian and a voiced dental fricative in Meglenoromanian and Romanian. Examples such as Romanian *ziuă*, Meglenoromanian *zuwă*, Aromanian *dzuwã* – all ultimately from Latin *dies* 'day' – illustrate the possibility of a parallel development of /dz/ > /z/ in Meglenoromanian and Romanian and a shared innovation of /i/ > /u/ before /w/ in Aromanian and Meglenoromanian. Moreover, /dz/ > /z/ occurs in the Gopeš-Molovište dialect of Aromanian in southwestern North Macedonia, where it is clearly an independent development (Wace & Thompson 1914: 251).

a transference of the ethnonym based on a profession or lifestyle commonly associated with an ethnic group. In Albanian, the opposite occurs, and *çoban* 'shepherd' comes to mean 'Vlah.' In Serbia and Bulgaria, the ethnonym *Vlah* is used to refer both to people from Wallachia (i.e., Romania south of the Carpathians) – and, by extension, Romania as a whole – and to Romance-speakers south of the Balkan range.[64] Moreover, in Serbia the term *Vlah* is also used to refer to Romanian speakers in eastern Serbia around Negotin and the Timok valley.[65] In this latter sense, it is an ethnographic term referring to a group that differed from Romanians not in language but in a specific set of historical circumstances that led to their settling in eastern Serbia during the Ottoman period. Former Yugoslav census figures classified these Romanian speakers as Vlahs together with the Aromanian-speaking Vlahs of North Macedonia, and so one must examine figures at the republic level for an accurate picture. While Aromanians themselves use the ethnonym *Armîn* (in the south) or related forms such as *Rămăn* (in the north), all etymologically from *Romanus* 'Roman' and historically involving loss of the short, unstressed /o/ and an elimination of the resulting /rm/ as an initial cluster, Meglenoromanians designate themselves with the Macedonian form *Vla* (plural *Vlaš*) in their own language. (They are also known in the local Macedonian dialect as *Pajakaški Vlasi* after Mt. Pajak, where most Meglenoromanian villages are located.) Here we use South Danubian Balkan Romance (SDBR) as a cover term for Aromanian and Meglenoromanian when they can be treated together. The term *Vlah* is used only when quoting from another source.

The term Macedo-Romanian has also been used to refer to Aromanian or Aromanian together with Meglenoromanian in opposition to the term Daco-Romanian, which refers to Romanian (or Romanian and Moldovan, depending on time and politics). On the politics of Dacian identity in Romania, see Verdery 1991; cf. Dietler 1994 on France and the Gauls. In both cases a nation-state – or

---

64 For discussion of other allonymic forms such as *Kutsovlah*, *Karavlah*, *Morlak*, *Beli Vlasi*, *Cincari*, etc., see Poghirc 1989: 9–11. We can observe here that the use of adjectives meaning 'black' (*kara*, *mavro-*) and white (*beli*) denote 'north' and 'west' [of the Danube], cf. the use of terms meaning 'black sea' and 'white sea' in Turkish and other languages to refer to the Euxine and Aegean, respectively (Poghirc 1989: 11). According to Athena Katsanevaki (p.c.), the term Κουτσοβλάχος in Greece is limited to Gramošteni (the Aromanians identified with Mt. Grammos). Historically, Aromanians have lived on the plains of Myzeqe, greater Epirus (i.e., both Çamëri and Ípiros), Thessaly, and geographic Macedonia, although some small groups may still be living in the Rhodopes. There is also a significant Aromanian population in Romania (Dobrudja, especially Constanța and Tulcea), but these are colonies that emigrated from the southern Balkans after World War One. Saramandu 1971: 1353 estimates the Romanian Aromanian population at 30,000, but elsewhere (1984: 423) he gives the figure 80,000–100,000, of which approximately 50,000 are in Dobrudja. We are not including here the Istro-Romanians, who constitute a separate case both historically and linguistically, and are considered to have branched off from Romanian (Todoran 1977: 106–107).

65 The designation of *Vlah* as an ethnicity for the Romanian speakers of eastern Serbia as distinct from the Romanian speakers of Vojvodina dates from 1946. In Romania, this designation is seen as an attempt to reduce the total number of Romanians in Yugoslavia (Lozovanu 2012ab). Speakers themselves are divided among those who identify as Serbian, those who identify as Romanian, and those who identify as Vlah. In recent decades, Vlah-identified speakers have devised an orthography and treated their speech as a separate language.

certain powers within it – that has a Romance official language has made strategic political use of the pre-Roman occupants. However, at present the term Macedo-Romanian implies that Aromanian is a dialect of Romanian (which, as a nation-state language, never takes a prefix, e.g., in published grammars of the language) rather than a separate language. On the other hand, it is worth noting that some Aromanian speakers in Romania refer to their language as *macedonean* 'Macedonian.' Istro-Romanian, spoken in Istria, left Romania sometime between the thirteenth and sixteenth centuries (Sărbi & Frăţilă 1998: 35–43, but see also Filipi 2002) and is also sometimes given separate status, but it is not normally treated in accounts of Balkan linguistics. In a sense, the status of Istro-Romanian is like that of Arbëresh within Albanic (Friedman 2001b).[66]

Also in our discussion of East Balkan Romance we must address the question of Moldavian/Moldovan. The standard language of the former province of Bessarabia (the territory of the former principality of Moldavia between the rivers Prut and Dniester, which was ceded by Turkey to Russia in 1812, declared independence in 1917, joined Romania in 1918, became a Soviet Socialist Republic in 1944–1947, and declared independence as the Republic of Moldova in 1991), known as Moldavian (subsequently Moldovan) is based not on Bessarabian dialects but on the same Wallachian dialects as Standard Romanian (see Dyer 1996). There has also been vacillation in Moldova over whether to call the language Moldovan or Romanian. Thus, aside from factors resulting from Russian or Ukrainian influence, Literary Moldavian was in fact not an elaborated separate dialect but rather a form of Literary Romanian written in the Cyrillic alphabet until 1989, when Latin became the official alphabet. In 1994, the official language of Moldova was Moldovan.[67] Since December 2013 it has been Romanian.

In terms of dialectal divisions, the main distinctions for Romanian proper are Moldavian (Moldova and Bucovina), Muntenian (Wallachia, southern Transylvania, and southern Dobrudja), Banat (including northeastern Oltenia), Crişana (including western Transylvania), and Maramureş (Caragiu-Marioţeanu et al. 1977). We can also mention here Beas (Boyash, Banjaš, etc.), a Romanian dialect spoken by groups of Romani descent (see also footnote 77) in Hungary, former Yugoslavia, Bulgaria and isolated groups in Greece and, perhaps, Albania, the archaisms of which indicate a separation from western (Banat, Crişana) dialects

---

66 Istro-Romanian is not to be confused with Istro-Romance or Istriote, which is considered to be a relative of Dalmatian or Venetian and, like Dalmatian, is poorly attested and died out in the nineteenth century.

67 Radio Free Europe/Radio Liberty Newsline (Vol. 2, No. 168, Part II, September 1, 1998) describes the name-change process thus: "Moldova on 31 August marks its *Limba noastră* ('Our Language') holiday, the country's second most important public holiday after Independence Day (27 August). Romanian was granted the status of 'state language' on 31 August 1989, but the parliament in 1994 changed the official state language to 'Moldovan.'" Then, in 2013 the Moldovan Constitutional Court ruled that the Declaration of Independence took precedence over the Constitution and that therefore the official language of Moldova was Romanian. The parliament passed a law to bring the Constitution into agreement with that ruling in 2023. Also, after World War Two, a strip of territory beyond the Dniester was added to Moldova. At present, this constitutes the unrecognized state of Transnistria, where Moldovan is co-official with Russian and Ukrainian and written in Cyrillic script except in some private schools. See Dyer 1996, 1999, 2022 for discussion on the status of Moldovan.

of Romanian in the eighteenth century (Sikimić 2005a, 2005b; on Albania see Weigand 1895: 78, see also Kahl 2012; Kahl & Nechiti 2012).[68] In the case of Aromanian, although there are many subdivisions, according to Saramandu (1984: 427), a basic distinction can be drawn between the dialects of the north – especially of the Fărşerot, and Grabovean, of southern Albania and adjacent parts of North Macedonia and Epirus as well as Gopeştean and Maloviştean of North Macedonia, which are characterized by absence of a distinction between schwa and high-back-unrounded vowel (<ă> vs. <î> or <â> in Romanian orthography) – and the dialects of the south, especially Grămostean and Pindean of Epirus, eastern Macedonia, and Thessaly, which distinguish schwa from the high back unrounded vowel (cf. Kahl 2005; Saramandu & Nevaci 2006, 2014a). Owing to patterns of migration over the past two centuries, followed by the hardening of borders in the twentieth century, which altered or eliminated the traditional patterns of transhumancy, these differences are realized in North Macedonia as a west/east opposition. The southern group as represented by the Grămostenii, are found east of the Vardar, while west of the Vardar is a mixture of various northern groups. Thus, for example, Saramandu 1984: 427 distinguishes Bela (near Struga) as well as Gopeš and Molovište (Bitola region) as distinct within the northern group, but Bela is actually divided into two groups: the older Măbalot (from *beala*, via *mbeala*, ultimately from Moscopole, Alb Voskopoja) and the more recently arrived Fărshălot (see also Friedman 1994b).

The poorly attested, extinct Eastern Romance language, Dalmatian, like Istro-Romanian, does not generally figure in accounts of Balkan linguistics. Dalmatian is known mainly from word lists from the last speaker, who was from the island of Krk (Itl Veglia, hence also the name Vegliote for Dalmatian) and died in an accident in 1898 (Fisher 1976; Maiden 2004 and references therein). On the other hand, the Venetian dialect of Italian exerted an influence on the western and southern coastal Balkan peninsula comparable to that of Turkish in the interior.

Finally, there is Judezmo, which is the language brought by Hispanic-speaking Jews of the Iberian peninsula to the Ottoman Empire (and elsewhere) when they were expelled from Spain on August 2, 1492 and Portugal in 1496–1497.[69] Coming

---

68 Autonyms such as Rudari and Lingurari are also used, as is the term Karavlah, which can also denote Romanian speakers not of Romani origin. Some groups do not regard themselves as Gypsies, and some are regarded as Gypsies by their neighbors but not by themselves. (Our use of *Gypsy* here translates local *Tsigan*, which constitutes a specific identity category – and sometime autonym – for some non-Romani-speaking groups of Romani origin; see also footnote 78.) For some, a non-Gypsy identity was crucial for surviving the Holocaust (Sikimić 2005a). Aromanians are sometimes called Cincar or Kutsovlah, terms which some speakers regard as pejorative.

69 Among the other names for this language are Ladino and Judeo-Spanish. Among specialists, the term Ladino is sometimes reserved for a written form of Judezmo that was used to translate Hebrew religious texts word-for-word. In using the name Judezmo, we follow Bunis 2018: 185–187, who gives a thorough discussion of almost all the names this language has been known by since its inception. (It can be added that the term Spanyol has been heard (so VAF, fieldnotes) in what is now North Macedonia.) Here the conclusion of Bunis 2018: 187, after his discussion of the various names, their origins and meanings, is worth quoting:

> Nevertheless, djudezmo still enjoys some popular use among native speakers and is the name preferred by many Jewish-language scholars – as a unique innovation arising within the speaker community; because of its designation of the language as a 'Jewish language',

to the Balkans at the invitation of Ottoman sultan Bayazid II, speakers of Balkan Judezmo spread out into mostly urban communities all around the eastern Mediterranean, in particular in Greece (especially the north and particularly in Thessaloniki and Kastoria, but also other towns (e.g., Véroia, Lárisa) and islands both in the Aegean Sea (e.g., Chios and Rhodes) and the Ionian Sea (e.g., Corfu), North Macedonia (especially Bitola, Skopje, and Štip but also Ohrid and elsewhere), Serbia, Bosnia-Hercegovina, Dalmatia, Bulgaria, and Turkey (most notably Istanbul and Izmir), and later (in the nineteenth century, coming mainly from Bosnia and Bulgaria) in Romania, too (particularly Bucharest). The language can be divided into two major dialects, Eastern and Western. Eastern Judezmo includes the dialects of Istanbul, Izmir, Rhodes, and Thessaloniki, while Western Judezmo includes the dialects of Belgrade, Sarajevo, Bitola, Bucharest, and Sofia. Most speakers of Judezmo were murdered in the Holocaust. Small remnant communities survive in Balkan towns (especially Istanbul and Thessaloniki but also elsewhere), Israel, and the United States. See Sala 1976 for a valuable and comprehensive bibliographic essay on the language. See also Quintana Rodríguez 2006, whose atlas provides a more complex and nuanced picture of various isoglosses. According to her, the northeast is more distinct from the west and southeast (Quintana Rodríguez 2006: 358).

1.2.3.4 Balkan Slavic

It is generally agreed that the Indo-European dialect that became Slavic acquired its recognizably Slavic shape in the northern part of Eastern Europe some time prior to the migration of Slavic speakers into the Balkans south of the Danube during the sixth and seventh centuries CE. On the basis of some old shared developments, South Slavic can be divided into two groups: East South Slavic and West South Slavic. The northernmost of the West South Slavic dialects became Slovene in what is today Slovenia and adjacent bits of Italy, Austria, and Hungary. There is linguistic evidence indicating that what became Slovene and what became Slovak (the southernmost of the West Slavic languages) remained in contact until the early tenth century CE, when Germanic and Hungarian speakers came between them approximately on the territory of today's Austria and Hungary. The more southerly and larger group that became speakers of West South Slavic probably consisted of a single people, most likely the *Slaveni*, who at some point were divided and ruled by two military aristocracies, probably of Iranian origin, who have been identified as the sources of the modern ethnonyms *Serb* and *Croat* (see Fine 1983: 53–57 for details). Their dialects became the West South Slavic that eventually occupied the territory of today's Croatia, Bosnia-Hercegovina, Montenegro, and Serbia (with later migrations to what became Italy, Austria,

sharing terminological parallels with some other Jewish languages (e.g., Yiddish); and as a memorial to major Judezmo-speaking communities, such as those of Salonika, Bitola (Manastir), and Rhodes, many of whose everyday members called their language djudezmo until they were annihilated in the Holocaust.

Hungary, Romania, and elsewhere). East South Slavic designates the Slavic dialects currently spoken on the territory of Bulgaria, North Macedonia, and adjacent parts of Greece, Albania, and southwesternmost Kosovo, but which we know to have been spoken all the way down to the tips of the Peloponnesian peninsulas, where a Slavic-speaking tribe known as Melingi were attested at least as late as the fifteenth century (Vasmer 1941: 18–19; Fine 1987: 166, 234). The evidence of toponymy also indicates the former presence of Slavic-speaking populations in other parts of what are today Albania and Greece (Vasmer 1941; Ylli 1997–2000).[70]

In connection with the conversion of the South and West Slavs to Christianity during the second half of the ninth century, the first known documents in a Slavic language were produced.[71] Byzantine missionaries from Thessaloniki, Methodius and his brother Constantine (who took the monastic name Cyril shortly before dying), are credited with having written or supervised the writing of these earliest documents, mostly translations of scriptures and other religious texts from the Greek, but also some translations from Latin and Old High German, and even some original compositions (Lunt 2001: 10). The originals have all been lost, and all that survive are later copies of a fraction of these texts, none of them older than the late tenth century, and most of them later. Based on the language of these copies, we can determine that in the late ninth century the various Slavic dialects, while already differentiated enough to be identifiable with certain regions, were nonetheless still mutually intelligible and not very far removed from what we can reconstruct as Late Common Slavic. The language of these earliest documents is called Old Church Slavonic, and it is defined as the non-East Slavic language of manuscripts and monuments that display certain archaic characteristics and are presumed or known to have been written prior to 1100 CE (see Lunt 2001: 1–14).[72] Documents dating from after this period and/or not preserving diagnostic archaisms are known as Church Slavonic (in various recensions) or as the "Old" stage of the various modern South Slavic languages.[73]

---

70 On the one hand, Serbian *Niš* and Macedonian *Štip* from Latin *Naissus* and *Astibo*, respectively, indicate transmission through an intermediary with what could have been Common Albanian phonological developments. On the other hand, Alb *Durrës* from Lat *Dyrrachium* indicates the intermediary of Slv *Drač* (from **dŭračĭ*), although the required initial stress is a problem (see Hamp 1966: 103–106). Further, modern *Ohrid* can be derived from Greek Λυχνιδός via a complex interaction between Slavic and Tosk Albanian (see Hamp 1981–1982).

71 The modern West Slavic languages are Czech, Slovak, Upper and Lower Sorbian, Polish, and Kashubian as well as the relatively recently deceased Slovincian and Polabian. Rusyn is now also a recognized distinct language, with a position transitional between West and East Slavic.

72 Among those distinguishing characteristics are the preservation of a distinction between the reflexes of Common Slavic **ŭ* and **ĭ*. Conservative scholars exclude from the OCS canon any document not displaying such distinctions, even if it can be determined that they are pre-twelfth century (Lunt 2001: 4–6).

73 It is the practice of Bulgarian scholars to include Old Church Slavonic with Old Bulgarian. Most of the earliest manuscripts that we have display evidence of distinctive phonological developments that are characteristic of modern Bulgarian dialects as well as adjacent dialects of modern Macedonian. Nonetheless, it is generally recognized today that the Old Church Slavonic tradition was to some extent a supradialectal norm with variants reflecting local usage, e.g., in regions where the dialects developed into Slovene or Czech (see especially Lunt 1982 on the problem of the dating

From the end of the Middle Ages to the beginning of the nineteenth century, the ancestors of what are now the modern South Slavic languages developed under different conditions in different regions and at various times. During these centuries, wars and population movements considerably complicated the dialectological picture of South Slavic. In terms of documentation, traditions varied from the vibrant vernacular literature of the Renaissance Dalmatian coast to the highly conservative but vernacular-influenced Church Slavonic ecclesiastical texts (*damaskini*) of Ottoman lands, the brief flowering of written colloquial language during the Reformation in Carinthia, Carniola, and Styria (of the fifty or so books published between 1550 and 1598, the Bible translation was especially significant), the mixture of Church Slavonic and Russian used by Orthodox Slavs in Hungarian lands, the vernacular literature composed by Slavic-speaking Muslims using Arabic script, and legal documents and chancery records of varied provenance, to name only some of the many types of written sources. In terms of the modern standard languages of today, however, in the Balkans, as in much of the rest of Europe, it was the rise of romantic nationalism and associated events and movements from the late eighteenth century onward that resulted in the current linguistic situation. Prior to that period, i.e., throughout the centuries during which the processes that led to the Balkan sprachbund were taking place, the names of speech forms and even the very concept of *language* were not isomorphic with our present concepts. (For an interesting discussion of the Western European developments that ultimately contributed to the construction of language that functions today in the Balkans and elsewhere see Bauman & Briggs 2003: 19–70.)

From a dialectological point of view, South Slavic linguistic territory presents an extraordinarily varied picture whose complexity can be captured only partially by the designation of isoglosses (cf. Alexander 2000a). Nonetheless, for heuristic purposes one can identify phonological and morphological developments whose geographic extent can be represented cartographically. The picture that emerges is a series of isoglosses that tend to cluster in certain areas but nevertheless do not define any radically sharp breaks (see Ivić 1958: 31, 32 and Friedman 1999a: 7). It is thus never the case that speakers from neighboring villages cannot understand one another, but as the differences increase with distance, eventually speakers from sufficiently separated villages will speak mutually unintelligible dialects, and those distances are less in some areas and greater in others. One group of isoglosses clusters in the region between what are today Slovenia and Croatia and another group clusters between the regions that today are located in southeastern Serbia and western Bulgaria and fans out across what is now the Republic of North Macedonia.[74] During the course of the nineteenth century, a variety of projects for South Slavic standard languages were pursued with varying degrees of success.

and provenance of OCS). As Lunt 2001: 4 points out, the term Old Macedonian would be just as applicable as Old Bulgarian. Old Church Slavonic is thus viewed not as the ancestor of any modern Slavic language but rather a kind of bachelor uncle on the South Slavic family tree.

74 As M. Greenberg 1994, 1996 has shown, even the transition from Slovene to Kajkavian Croatian accent systems proceeds by regular increments, without missing links.

The details of these processes need not concern us here.[75] As of this writing there are seven South Slavic official languages (moving roughly from northwest to southeast): Slovene, Croatian, Bosnian, Montenegrin, Serbian, Macedonian, and Bulgarian. There have been a number of attempts to develop standards from various other dialectal bases in all of these languages over the years, but for simplicity's sake, we limit ourselves mostly to languages with titular nation-state support (but see below for some specific exceptions). Our task here is to determine how these relate to our object of study, the Balkan sprachbund.

Since the Balkan sprachbund, like all such contact phenomena, is the product of orality, the linguistic level with which we are concerned is the dialect. At the same time, however, space precludes the exhaustive exposition of every dialectal datum. To the extent that they are based on speech, standard languages provide a convenient tool for summarizing the relevant phenomena. South Slavic, however, is the only language group with contiguous dialects some of which are clearly Balkan and others of which clearly are not. The question therefore arises as to where to draw the line. In §1.1, we indicated competing definitions and claims relating to the borderland Balkan Slavic dialects as well as some of the terminological problems. In this book we refer to the Slavic dialects of Bulgaria and of Greek and Turkish Thrace as Bulgarian, the Slavic dialects of the Republic of North Macedonia, Greek Macedonia, and adjacent parts of Albania as Macedonian, and the Slavic dialects of the Republic of Serbia as Serbian or BCMS. The following exceptions are hereby noted: The dialects of the Bosilegrad and Dimitrovgrad (Caribrod) districts of the Republic of Serbia are acknowledged by both Serbia and Bulgaria as Bulgarian and the speakers consider themselves to be ethnic Bulgarians. The dialects of the Slavic-speaking Muslims of Gora, a region on the eastern and northern slopes of Mts. Korab and Šar in northeastern Albania and the adjacent southwesternmost corner of Kosovo are included with Macedonian unless otherwise specified.[76] Macedonian-speaking Muslims are sometimes known as *Torbeš* (Alb *Torbesh*), a term that is also applied to Goran in North Macedonia and Albania. Some Muslim Macedonian speakers consider this term derogatory, while others embrace it. Serbian dialects in the Republic of North Macedonia are of

---

75 See the articles in Schenker & Stankiewicz 1980 and Picchio & Goldblatt 1984 for discussions.

76 Until the 1980s, the dialects of Gora were categorized as Serbian by Yugoslav linguists. In the 1980s, however, Vidoeski 1986 argued on the basis of such phonological criteria as fixed antepenultimate stress that the dialects were better classified as Macedonian, and these arguments were already accepted in Ivić 1985. In Bulgarian dialectology, Kočev 2001 differs from Kočev 1988 in that the older of the two maps accepts the earlier classification of the dialects of Gora as Serbian and therefore excludes them from Bulgarian claims, while the more recent map accepts the reclassification of those dialects as closer to Macedonian and therefore includes them in Bulgarian claims. In Albania, Gorans have a distinct identity, which, however, they are somewhat reluctant to reveal owing to political pressure. In Kosovo, some Gorans also have a distinct identity, some identify as Macedonian, some as Serbian, and some as Bosnian (this last on the basis of the claims of Bosnian Muslims to include all Slavic-speaking Muslims of former Yugoslavia); still others identify or identified as Turks on the basis of religion (see Nomachi 2018). With regard to the South Slavic standard languages, Gorans in Albania are aware that their dialects are closer to Macedonian than to Serbian (VAF field notes 1995), and some identify their language as Macedonian, although others, having been paid by Bulgarian agents, identify it as Bulgarian (VAF field notes).

two types. The dialect of the Gallipoli Serbs of Pehčevo near Berovo represents a Serbian enclave settled in North Macedonia from Turkish Thrace in the wake of the Balkan Wars, World War One, and the subsequent Greco-Turkish War (Ivić 1957: 5–6). Certain dialects in the north of the Republic of North Macedonia such as the village of Kučevište and other villages on Kozjak and Skopska Crna Gora north of Skopje are considered by their speakers to be Serbian because they identify as Serbian (Orthodox) and learn standard Serbian in school. These dialects, however, do not differ from the Macedonian dialects of the neighboring villages in terms of any structural features (Vidoeski 1998: 10). The dialects of Bulgarian-speaking Muslims can be specified by the term Pomak. At present Pomak is treated as a separate minority language in Greece (Kokkas 2004; Karakhotza 2006; Papadēmētriou 2013; Theocharidis 1996a, 1996b).[77] For Bulgarian dialectologists, these Pomak dialects are part of the same Rhodopian dialect complex of Bulgarian spoken on both sides of the mountains/border (cf. Steinke & Voss 2007; Friedman 2012c). While there are some ethnolectal differences between the dialects of Christians and Muslims in the region (cf. R. Greenberg 1996a), the Slavic speakers of the Pirin Macedonia region (Blagoevgrad district of Bulgaria) are divided in terms of self-identification: some are Bulgarian-identified and some are Macedonian-identified. The Slavic dialects of Pirin Macedonia are therefore specified here as such. There are also Bulgarian enclaves in Romania and Moldova as well as Slavic-speaking Muslims in Turkey. We can also note here the presence of Serbian and Montenegrin dialects in northern Albania and a Bosnian enclave near Durrës and another enclave near Fier (Steinke & Ylli 2013).

1.2.3.5 Balkan Indic (Romani)

Like other modern Indic languages, Romani developed as a distinct language or dialect in India during the Middle Indic period, defined as 600 BCE–1000 CE (Masica 1991: 51), probably in Central India (see Matras 2002: 14–18, for a summary of the evidence). This means that in terms of attestations of earlier stages we have at our disposal the same data as for the rest of Indic: Vedic, Classical Sanskrit, Prakrits, and Apabhraṃśa. There is, however, a significant gap between these ancient languages and our first Romani documents, the oldest three of which are short lists of words and phrases dating from the sixteenth century (1542, pre-1570, 1597), all from England and Holland (Friedman & Dankoff 1991: 1; Matras 2002: 10). The fourth oldest Romani document, however, is from the Balkans, namely Komotini (Trk Gümülcine), in what is now Greek Thrace. It is a list of words and phrases with Ottoman Turkish translations in Evliya Çelebi's travelogue, the *Seyāḥat-nāme* (Friedman & Dankoff 1991). Based on linguistic evidence, it seems likely that Romani was in contact with Byzantine Greek in Anatolia by

---

77 See also Greek IV Army Corps (1995–1998). The IV Army Corps, with garrison headquarters in Xánthi, is responsible for guarding the Greco-Turkish border, which is the river Evros (Trk Meriç, Blg Marica).

1000 CE (Tzitzilis 2007b), although we cannot say for certain when the Roms crossed from Asia Minor to the Balkans other than to note that they were already in Constantinople by the middle of the eleventh century (Soulis 1961: 144).[78] By the fourteenth century they had already begun their migrations to other parts of Europe.

Although Romani attracted the interest of some of the earliest Balkanists (Miklosich 1874–1878; Weigand 1895: 78), it was mentioned only to be excluded during the "classical" (or modern) period of Balkan linguistics. Thus, for example, Sandfeld 1930: 3 mentions Romani in a footnote without any attempt to integrate it into his findings, and Asenova 2002: 220 describes Romani as *edin nebalkanski ezik, no s dălgo žitelstvo na Balkanite* 'a non-Balkan language but with longtime residence in the Balkans.' Some individual studies of Balkan features and the Balkanization of Romani were published during the late twentieth century (Kostov 1973; Friedman 1985a; Matras 1994a), and Joseph 1983a does spend some, albeit limited, time on Romani in his monograph on the Balkan infinitive, but Romani has only recently begun to be integrated into any more extensive study of the Balkan languages. Thus, for example, Sawicka 1997 takes Romani into consideration, and Hinrichs 1999a contains two chapters treating Romani in its Balkan context (by Boretzky & Igla, and by Bochmann). This is in part owing to the general marginalization of Roms and in part to the lack of both adequate data and national structures for Romani. Among the reasons Romani provides important contrast to the rest of the Balkan languages is the unidirectionality of ordinary multilingualism, i.e., Roms learn other Balkan languages, but most non-Roms do not learn Romani. The same could be said about Judezmo.[79]

At present, the dialectological classification of Romani accepted by most linguists working on the language distinguishes four main groups: Balkan, Vlax, Central, and Northern (cf. Matras 2002: 5–13, 214–237). Each of these is subdivided into two groups apiece: Balkan I and II, Northern and Southern Vlax, North and South Central, and Northeastern/Northwestern.[80] The two types of Balkan dialects are spoken by groups that have stayed, for the most part, in the so-called

---

78 Many English-language writers use the Romani substantival plural form *Roma* as a singular and plural noun and also as an adjective. As noted elsewhere (Friedman & Hancock 1995), however, just as in English the plural of *Turk* is *Turks* and not (the Turkish form) *Türkler*, so the plural of *Rom* should be *Roms* rather than *Roma*. The form *Roma* exoticizes and marginalizes rather than emphasizing the fact that the group in question is an ethnic group just as are Turks, Magyars (not *Magyarok*), Bulgars (not *Bulgari*), etc. We note that in all the Balkan languages, the ethnonym *Rom*, when it is used, takes those languages' normal plural, e.g., Bulgarian, Macedonian, Romanian, Serbo-Croatian *Romi*, Albanian *Romë*, etc. The use of *Roma* as an adjective is a solecism. The adjective, in English (which does not inflect adjectives), is *Romani* (cf. *Hindi, Nepali, Punjabi,* etc.). The spelling *Romany* is an archaism on the level of *Hindoo* for *Hindu.* On occasion we use the term Gypsy as an ethnonym for groups that do not speak Romani albeit being of Romani origin (as in §1.2.3.3) or as a translation or citation form where appropriate. See also footnote 68.

79 So also with respect to scholarship on Judezmo: very few Balkanological studies treat the language or include it in surveys; Afendras 1968 seems to be one of the earliest works in the modern period of Balkan linguistic studies to treat developments in Judezmo as relevant to the examination of the other Balkan languages.

80 Certain individual dialects do not fit easily into this classification scheme and must be considered as transitional or isolated in one way or another, but none of these need concern us here (see Matras 2002: 236–237).

Southern Balkans (Albania, Kosovo, North Macedonia, Bulgaria, Greece, and Turkey) but also by groups that have migrated to Romania, Crimea, Azerbaijan, and Iran (Matras 2002: 6). The Balkan I dialects are more conservative and are spread throughout the area, while Balkan II are found in northern Bulgaria, Kosovo, and adjacent parts of North Macedonia and Albania. Of the former, the Arli (from Turkish *yerli* 'local') dialect is particularly widespread while of the latter, Bugurdži (from Turkish *bürgücü* 'gimlet-maker' also called *kovači* 'blacksmiths' in North Macedonia, as well as other names) is the predominant variety spoken in North Macedonia and Kosovo (Boretzky 1993, 2000b; Matras 2002: 223). The Vlax group, so named because of significant Romanian lexical influence presumed to have been acquired during an extended sojourn on Romanian-speaking territory, when certain shared phonological innovations developed, is divided into a Southern branch, which migrated back into the southern Balkans (*Gurbet* 'migrant,' *Džambaz* 'horse-dealer,' etc.), and a northern branch (*Kalderaš* 'kettle-maker,' *Lovari* 'horse-dealer,' etc.), some of whom remained in Romania, others of whom migrated south to northern Bulgaria, west to central Serbia, north to Hungary, Poland, and beyond, and many of whom joined the Eastern European emigration to North America in the late nineteenth and early twentieth centuries. The Central group is concentrated in former Austria-Hungary, the Northeastern in Russia, the Baltic lands, and central Poland, and the Northwestern in Germany, France, the Nordic countries, and, into the twentieth century, Great Britain (see especially Matras 2010).

The Balkan versus non-Balkan varieties of Romani are of particular interest also because Romani as such acquired its basic shape during its sojourn in the Balkans, and subsequent migrations have resulted in a differentiation between those dialects spoken in the Balkans, which continued to develop Balkanisms in contact with the other Balkan languages (e.g., future formation), and those dialects spoken outside the Balkans, which lack some basic Balkan features.

1.2.3.6 Balkan Turkic

The oldest Turkic monuments, the Orkhon inscriptions of the upper Yenisei, date from the eighth century CE and show remarkable affinities with Oghuz Turkic (Tekin 1968). As noted above (§1.2.1.11), we are interested here not only in Turkish as an adstratum, which has been the traditional approach, but also in Turkish as a participant in Balkan linguistic processes. That said, we can observe that although various Turkic-speaking peoples passed through or settled in the Balkans as noted in §1.2.1.11, we are concerned in this section only with the arrival of the speakers of the dialects that became Balkan Turkish in the Balkans. These dialects can be traced to the arrival of Oghuz-Turkic speakers in Anatolia in the eleventh century, a convenient date being the defeat of the Byzantine Emperor Romanus IV Diogenes by the Seljuk Turks under Alp Arslan at the Battle of Manzikert (Trk Malazgirt) in 1071, which opened Anatolia to Turkic conquest. Although – under pressure from the Kumans (Polovtsy) – a group of Ghuzz (Oghuz) Turks invaded the eastern Balkans from the north in 1064, most of them were wiped out

by the plague and the rest scattered or became Byzantine mercenaries (Fine 1983: 211). As Fine 1987: 165 observes, by 1261, when the Byzantines recaptured Constantinople from the crusading Latins ("Franks"), who had held it since 1204, "Byzantium was hardly an empire any longer – despite its titles, rhetoric, and court ceremonial; it was just another petty state, holding, together with Constantinople, western Anatolia, Thrace, Thessaloniki, and Macedonia." This set the stage for Ottoman expansion during the following century. By 1300, most of Anatolia was in Turkish hands, with Osman I, who established the Ottoman dynasty, ruling an emirate in northwestern Anatolia 1290– 1326. During the mid-fourteenth century, various Turkish troops were used in Europe by rival Byzantine dynasties, and in 1352, a Turkish army defeated a Serbian one – each fighting for opposite sides in a Byzantine civil war – at Demotika (Fine 1987: 325–326). Technically, this was the first major Turkish–European battle in Europe. During this period, however, Turks in Europe were raiders and mercenaries rather than settlers. In 1354, taking advantage of an earthquake that had collapsed the walls of Gallipoli, the Turks crossed the Dardanelles and occupied the fortress for themselves, an event which marked the beginning of their occupation of Europe as a political force. Adrianople (modern Edirne) fell in 1369 (Fine 1987: 406). The decisive defeat of Serbian forces occurred at Chernomen (Grk Orménio) on the River Marica in 1371 (Fine 1987: 379) – near what is today the Turkish–Greek–Bulgarian border – although the later Serbian defeat at Kosovo Polje in 1389 is more famous. By the end of the fourteenth century Ottoman rule covered all of what would become Balkan Slavic linguistic territory as well as what is today eastern Greece and southern Romania, and by the end of the fifteenth century Ottoman rule had expanded to include the entire region with which we are concerned here.[81] The core of the Balkan linguistic area as we have defined it remained under Turkish rule until the early twentieth century, although the peripheries began to assert their independence in the nineteenth. Even after the retreat of the Turks to eastern Thrace, however, Turkish remained a language of urban sophistication and prestige, especially in those countries that did not expel their Muslim populations (i.e., it remained so everywhere except in Greece).[82] Turkish was still spoken by town dwellers in North Macedonia regardless of religion well into the second half of the twentieth century (VAF, field notes,1973–2001) and it remains an important language for Muslims throughout the Balkans (cf. Ellis 2003 on North Macedonia).[83]

---

81 The fall of Constantinople in 1453 was an important symbolic and cultural victory for the Ottomans, and had a significant effect on the organization of the Ottoman state, but from the point of view of territorial conquest it did not add significantly to Ottoman lands. The Ottoman occupation of Wallachia, Moldavia (including Bessarabia), and Transylvania was complex. Wallachia became tributary to Bayezid I (1391), although there was occasional resistance until the death of Vlad Ţepeş in 1476. Moldavia was paying tribute in 1456 and submitted fully to Bayezid II in 1512. These vassalages retained the right to elect their own rulers until 1714, after which control passed to the Phanariotes until c.1830. Transylvania was conquered with (the rest of) Hungary by Süleyman I (Battle of Mohács, 1526, formal tribute 1556), and retained until the Treaty of Sremski Karlovci (Karlovitz, Karlóca) in 1699 (Sugar 1977).

82 On Bulgaria just after independence, see for example Herbert 1906: 152. Also, Turkish speakers were allowed to remain in Greek Thrace.

83 As late as the 1970s, US NATO troops stationed in Turkey could use Turkish when on vacation in the older sections of Athens (VAF field notes, 1979).

The Turkish dialects of the Balkans are divided into two groups, East Rumelian and West Rumelian (Németh 1956). The location of the boundary between the two is remarkably similar to that of the so-called *jat*-line (see Stojkov 1968: 56), a major isogloss distinguishing East Bulgarian from West Bulgarian (Hazai 1961). This is to say that the local Turkish dialects of Kosovo, Macedonia (including Greek Macedonia before 1923; cf. Mollova 1960), and Albania, as well as western Bulgaria (Kakuk 1960), are all of the West Rumelian type, except for the dialect of the Yürüks of eastern North Macedonia, who are more recent arrivals and speak an East type of Rumelian Turkish (Nedkov 1986; Jašar-Nasteva 1986; Manević 1953/1954). East Rumelian is basically an extension of Istanbul Turkish, while West Rumelian shows considerably more Balkanization at all levels of its grammar (Ibrahimi 1982; Friedman 2002a; cf. Katona 1969; Kowalski 1926). Gagauz has been classed as an East Rumelian dialect with Indo-Europeanized (i.e., Balkanized) syntax (Menz 1999), and in this respect is also of interest here.

1.2.3.7 Language Choice and Dialect versus Standard

Owing to the fact that the different members of a given linguistic group frequently display the same Balkanisms/Balkan linguistic phenomena (*ceteris paribus* and *mutatis mutandis*), it is generally the practice to cite one or at most two representatives from any given group, usually in standard orthography, with dialectal examples supplied only where particularly relevant. Sandfeld generally uses dialectal examples, in part because he was relying on folklore texts or on works published before the relevant languages had been standardized, and in part because some of the languages about which he was writing had not yet been standardized at the time he was writing.[84] All of the handbooks of the late twentieth and early twenty-first century have followed the practice of citing mostly standard forms taking Bulgarian as representative of Balkan Slavic and Romanian as representative of Balkan Romance. While such a practice is not in and of itself misrepresentative, as long as appropriate data from other languages and various dialects are cited where relevant, in this book we nonetheless follow a different practice based on the centrality of Macedonia for Balkan sprachbund phenomena in general. (For a discussion of Macedonia as the "heart" of the Balkan sprachbund, see Hamp 1977a; Topolińska 2010.) Except where the specificity of the data requires it, we take Standard Macedonian as our representative of Balkan Slavic, Standard Albanian as our representative of Albanic, and Aromanian as our representative of Balkan Romance. Since Standard Macedonian is based on its west central dialects and Standard Albanian is based on northern Tosk, the

---

84 Thus, for example, although Elbasan Geg was an official standard in Albania between the two World Wars, in reality it functioned as a kind of *lingua communis* (cf. Naylor 1992), and modern Standard Albanian (based on Korça Tosk), like Standard Macedonian, evolved in the wake of World War Two. It is worth noting in this regard that Sandfeld often distinguished between Bulgarian and what he called *macédo-bulgare* (cf. Friedman 1997c), despite the fact that at the time he was writing, the status of Macedonian was disputed (see Friedman 2000e; Vaillant 1938).

dialectal bases of the respective standard languages are at the heart of the heartland. Moreover, Aromanian represents the form of Balkan Romance with which the other languages have been in contact for the longest time. Since the standardization of Aromanian is still in progress (see Friedman 2001b), we take Gołąb's 1984a grammar of the Kruševo (Aro Crushuva) dialect and Markovikj's 2007 monograph on the dialects of the Ohrid-Struga region as basic (these dialects being among those that have been in contact with the dialectal bases of both Standard Macedonian and Standard Albanian). For Greek, we use the Demotic standard but with reference to northern or other dialects as appropriate. For Romani, we take the Arli dialects of Skopje as our base, these being both a particularly representative dialect and the base of standard Romani as used in the Republic of North Macedonia (the only country in the world to mention the Romani people (*Romi, romskiot narod*) in its constitution).[85] For Turkish, we use standard or West Rumelian as appropriate.

1.3   On Maps and Toponyms

As Wilkinson 1951 makes clear, ethnolinguistic maps reflect and participate in various types of political projects (see also Hertslet 1891). Furthermore, a map that attempts to capture the genuine multilingual complexity of the Balkans is basically unreadable (cf. Friedman 2007a). Recent maps are no different from those of a century or more ago. They focus on this or that historical moment to justify the boundaries they draw, or they select criteria that favor a simple or hegemonistic point of view. The following examples are illustrative. Since the beginning of the twentieth century, numerous maps have been published that claim to show "historical and ethnic Albania." In most cases, the boundaries of these maps are those of the Ottoman vilayets of Işkodra (Alb Shkodra), Yanya (Grk Ioánnina), Manastir (Mac Bitola), and Kosova (at various times Niš, Prizren, Prishtina, Üsküb [Mac Skopje]) (Karpat 1985).[86] Similar maps of "historical and ethnic Macedonia" use the *vilayets* of Üsküb (Kosova), Manastir, and Selânik with an extra *kaza* here and there.[87] "Ethnic Bulgaria" conforms mostly to the

---

85 In the original preamble and Article 78 (November 17, 1991) and in Amendments IV and XII (November 16, 2001). There are seven sub-varieties of Arli traditionally spoken in Skopje, of which Topaanli, the old urban dialect, and Gavutno, the dialect of surrounding villages, are the most differentiated, with other sub-dialects generally sharing various features with one or the other (see Friedman 2017a and references therein for details).

86 It is interesting to note that in 1999, when pro-Albanian circles were courting western public opinion and when thousands of Albanians were working in Greece as economic migrants, these maps were modified so that the claimed ethnic border coincided exactly with the Albanian–Greek political border (which was not the case in maps based on Ottoman boundaries).

87 As indicated in §§1.2.2.7.1 and 1.2.3.3, *kaza* is roughly 'county' and *nahiye* is roughly 'township.' Ottoman administrative divisions in the Balkans underwent numerous changes, especially in the course of the nineteenth century. The largest administrative unit was changed from the *eyalet* to the *vilayet* in 1864. The *vilayet*s were divided into *sancak*s (also spelled *sandžak, sandjak, sanjak*), with *kaza* and *nahiye* being smaller administrative subdivisions.

boundaries drawn at San Stefano (Yeşilköy) in March 1878 at the end of the Russo-Turkish War, although the ethnic claims differ in a few details. Nicolaïdes' ethnographic map of 1899 purported to define territory by "commercial language" or schools and had Greek extending more or less to the current political boundary of the Greek state even though at that time the linguistic boundary of Greek was much farther to the south. Taking certain isoglosses as diagnostic, Serbian and Bulgarian scholars have extended their territories to overlap with one another. In the Serbian and Bulgarian cases, Macedonia is completely erased as an entity. In contrasting Albanian with Macedonian presentations, there is overlap for the vilayet of Manastir and those sandjaks of Kosova south of Mount Šar and Skopska Crna Gora (Trk Karadağ, Alb Mal i Zi).

There is a certain commonality between Karl Sax's 1878 map and the 1994 Macedonian census map which is the frontispiece of this volume (and see below). According to Wilkinson 1951: 77, Sax's was the first map that attempted to combine linguistic and religious criteria. Sax distinguished a sense of community that he called *nationales Bewusstsein* 'national awareness/consciousness,' which Wilkinson implies was not based on race or folklore and was also something other than language, religion, or a combination of the two. His examples, however, all involve precisely combinations of religion and language, albeit not necessarily recognized as such by Sax (or Wilkinson): e.g., in today's terms, Bosniacs and Macedonians, whom Sax labels *Muslim Serbo-Croats* or *Bosnian Turks* and *Serbo-Bulgarians of Greek Orthodox religion*. Wilkinson 1951: 81 writes that Sax made the point that there were so many different nationalities in Turkey in Europe (p. 28) and with such complex intermingling that "no possibility existed of granting political independence to each group. Macedonia in particular had a very heterogeneous population … ." Sax's methodology "accentuated the confusion of nationalities" and was "related to Austrian policy in the Balkans in so much as it attempted to belittle the political significance of ethnic groupings." It is worth noting that ethnographic maps of Austria–Hungary produced after the Empire's occupation of Bosnia–Hercegovina in 1878 erased both religion and non-Slavic languages in the region it had occupied.

The map reproduced as the frontispiece to this volume (with permission) was produced by the Bureau of Statistics of what is now the Republic of North Macedonia during the extraordinary 1994 census. The map was based on answers to question #12 on form P-1, *izjasnuvanje po nacionalna pripadnost* 'declaration of national affiliation' and, in its full color form, combines seven colors with six schematic representations of a person. (See www.cambridge.org/BalkanLanguages for the color version.) The colors are blue (Macedonian), green (Albanian), red (Turk), yellow (Rom), orange (Vlah), brown (Serb), and white (other). The figures vary in size with each size gradation representing a power of ten or a multiple of five (10, 100, 500, 1,000, 10,000, 50,000). The figures are grouped on the map inside the 1994 boundaries of the thirty municipalities. They are arranged in rows according to size (large above small) but not according to color. Thus, for example, the second row of figures for the municipality of Struga, representing one thousand

people each, alternates red-blue-green-blue-green-blue-green-blue-red. For the top row for the municipality of Debar, a green figure representing 10,000 is sandwiched between red and blue figures of 5,000 on the viewer's left and red, blue, green, and yellow figures of 1,000 on the viewer's right. Moreover, although a key is provided, the visual effect is extremely difficult to interpret without a census table. The overall effect of the map is one of extreme intermingling and the sense that interpretation can only be achieved with hours of effort. This effect is arguably related to the kinds of motivations that can be imputed to Sax's map insofar as the point is that complexity does not justify fragmentation. In the context of 1994, the point of the map was that an ethnic partition of what was then the Republic of Macedonia would be an impossible project without doing violence to everyday lived complexity.

Thus, in any map of the Balkans, one must ask "when is the map?" and "why/for whom is the map?" If it is of the ancient Balkans, Dardania might appear north or south of Mount Šar (see Snively 2017), Illyria might also have wildly different boundaries (compare Gjinari 2007: Map Ç with Shepherd 1964: 10, 13), and so would other ancient regions. Maps of the medieval period also changed depending on political vicissitudes (see for example Fine 1983: 93, 244 and Ransohoff 2017) and depending on the century. For instance, Skopje was the capital of the Serbian empire, a major town in the Bulgarian Empire, and important center of Samuil's medieval kingdom, which is claimed by both Bulgaria and North Macedonia as ancestral, and also a town in the Byzantine Empire. During the five centuries of Ottoman rule, not only did the territory controlled by the Turks wax and wane, but the internal administrative boundaries also underwent numerous alterations, especially in the nineteenth century. Different groups choose different moments in the nineteenth century as the "historical moment" to their best advantage. While the complexities described above require a detailed and nuanced reading, choosing a criterion such as "50% or more of the ethnic group" (National Geographic 1999) – even if the percentage itself were accurate – gives an impression of homogeneity that does not occur in real life.

Therefore, in this book we do not attempt to give comprehensive historical or ethnolinguistic maps of the Balkans. The former are the task of an historical atlas: Shepherd 1964; Magocsi 1993; Hupchik & Cox 2001 are contributions (cf. also Wilkinson 1951; Crampton & Crampton 1996; and Cattaruzza & Sintès 2012; Darques 2017 provides a modern geographic approach). Ethnographic maps can only create a false impression if easy to read or a sense of hopeless confusion if they attempt to adhere as strictly as possible to the lived reality of the region. For the locations of the numerous toponyms referred to in this book, the ready availability of online resources such as Google Maps makes it easy for the reader to locate even the smallest hamlet, and a complete set of maps pertaining to everything described in this book would be an atlas unto itself.

As is the case in any multilingual region, most toponyms in the Balkans have different names in the various languages of the peoples that have occasion to refer to them. In some cases the differences are rooted in the phonological history and structure of the respective languages, e.g., Mac/Blg *Skopje*, Srb *Skoplje*, Alb *Shkup*,

Trk *Üsküp*, Aro *Scopia*, and Grk *Skópia*, all ultimately from Lat *Scupi*, itself of pre-Roman origin. Similarly, Grk *Thessaloníki* is called *Solun* in various Slavic languages, *Sãrunã* in Aromanian, *Selânik* in Turkish, *Selanik* in Albanian, and *Salonica* or *Salonika* in older English-language sources. In other cases, the name is quite different due to translation or calquing, as in the case of Slv *Crna Gora*, Alb *Mal i Zi*, Trk *Karadağ* 'Black Mountain' (Eng *Montenegro*, itself from Italian), similarly Mac *Bitola* and Aro *Bitulji* (from Slv *obitělь* 'monastery') but Trk/Alb *Manastir*, from (Mod)Grk *Monastíri* ('monastery,' cf. a similar phenomenon in Cro *Rijeka*, Itl *Fiume*, both meaning 'river'). Sometimes the toponyms have separate histories as in Mac *Veles* (tentatively identified with a pre-Christian Slavic deity) and Trk *Köprülü* 'having a bridge' (cf. *Mostar*, based on Slv *most* 'bridge,' in Hercegovina), or Srb *Metohija* 'monastery lands' (referring to the properties of the Serbian Orthodox Church), which in Albanian is *Rrafshi i Dukagjinit* 'Plain of Dukagjin' (referring to a medieval Albanian ruler and tribal grouping). Similarly Mac *Tetovo* (from older *Htetovo*) and Alb *Tetova* correspond to Trk *Kalkandelen*. An added complication is the fact that during the twentieth century various nation states and other entities engaged in conscious processes of toponym-changing for ideological or ethnopolitical reasons. Thus, for example, in the southwest corner of Bulgaria (the region of Pirin Macedonia), the town of *Gorna Džumaja* ('Upper Mosque') was changed to *Blagoevgrad* (in honor of Dimitar Blagoev, a pre-World War Two Marxist activist) in 1950 and *Nevrokop* was changed to *Goce Delčev* in 1951 after a revolutionary of the late Ottoman period claimed now by both Bulgaria and North Macedonia; in Greece, all the Slavic and Turkish village names in Greek Macedonia and Thrace have been replaced or Hellenized, e.g., *Ziljahovo/Néa Zíkhni*, *Nestram*/*Nestório*, *Smrdeš/Krystallopigi*, *Dedeağaç* (Pomak and pre-1926 English *Dedeagach*)/*Alexandroúpoli*, *Skeča*/*İskeçe*/*Xánthi*, etc.; in Turkish Thrace, *Kırk Kilise* 'forty church[es]' – in Greek *Saránta Ekklisies*, also 'forty churches,' but in Bulgarian *Lozengrad* 'vine town') – is now *Kırklareli* 'land of forties.' A recent example has been the replacement of etymologically Slavic toponyms in Kosovo in Kosovar Albanian-language publications since the late 1990s, e.g., *Srbica* (Alb *Sërbicë*) > *Skenderaj* (the former referring to 'Serb,' the latter to Skanderbeg, the medieval Albanian hero who defied Turkish rule; but see Schmitt 2009), *Suvareka* (Alb *Suharekë*) > *Theranda* (the former from Slavic for 'dry river,' the latter a pre-Roman toponym).[88] Note also the effects of Katharevousa and diglossia in Greece, e.g., the Dimotiki spelling *Livad(e)ia* corresponds to Katharevousa *Levád(e)ia* (and the Dimotiki form is a feminine singular while the Katharevousa form is a neuter plural). Choice of toponymic reference is an extremely sensitive issue, often connected with perceived claims of sovereignty, threats to territorial security, or some form of cultural hegemony (Friedman 1993a: 82–83). At the same time, differences of usage both in sources dating from various historical periods and in works written in the Balkan languages themselves necessitate that the interested reader be aware of the toponymic

---

88 But note that the classical Theranda was probably modern Prizren (see Hamp 1992a).

correspondences. In this book, our usual practice is to cite a toponym in its standard English form if one exists – e.g., *Greece* rather than *Ellas* or *Ellada* (corresponding respectively to the Katharevousa and Dimotiki forms), or *Athens* rather than *Athínai* or *Athína* (again corresponding respectively to the Katharevousa and Dimotiki forms) – or the current form of the majority language of the nation-state in which it is currently located, e.g., *Kırklareli*, not *Kırk Kilise*, *Saránta Ekklisies*, or *Lozengrad*.[89] If a given toponym has a significantly different form in the language or time period relevant to the immediate discussion, that form generally is given in parentheses. In some cases, however, the minority language name is the primary referent, and the current nation-state majority language name is given in parentheses. Such is the case, for example, with Meglenoromanian-speaking villages or when discussing Macedonian dialects spoken in Greece.

One toponym in this book requires a paragraph of its own: Macedonia. Of all the toponyms in the Balkans, none has been so contested as this one in the recent past, nor as changeable in its reference of the distant past.[90] In this book, we use unmodified *Macedonia* to refer to the region recognized more or less as such since Ptolemy and Strabo, and, *de facto*, by Wilkinson (1951: 3) and which, from the late fourteenth century until 1912–1913, was entirely within the Ottoman Empire. North Macedonia refers to the internationally recognized republic of that name. Pirin Macedonia refers to the southwestern corner of Bulgaria, mainly the Blagoevgrad district, and the terms Greek Macedonia and Aegean Macedonia both refer to the Greek territory between Epirus and Thrace and north of Thessaly. The term Aegean Macedonia is usually favored when the discussion concerns non-Hellenic languages spoken in what is now Greek Macedonia. Although this term is perceived as irredentist by Greek nationalists, as are all Macedonian toponyms in Greek Macedonia, this term, like the Macedonian toponyms in Greece, is used by Macedonian speakers themselves who are from or still live in the region. For this reason, we respect the usage of the speakers. Some Macedonian dialects are located in today's Albania and Kosovo (Prespa, Debar, Gora), but there is no collective term for the enclaves (see §1.2.3.4 for additional material).

1.4   Writing Systems

The Balkan languages have been written using a variety of alphabets; here we survey what is found for each language. For the specifics of the orthographies used in this book, the reader is referred to the User's Guide.

---

89 In the case of Albanian, there is considerable variation in using definite or indefinite forms, e.g., *Tirana* vs. *Tiranë*, *Elbasani* vs. *Elbasan*. We follow the common English-language practice of using the definite for feminine toponyms in -*a* (indefinite -*ë*) and the indefinite elsewhere. In the case of Kosovo, we use the form that is now internationally recognized.

90 The lone exception being, perhaps, Romania, which at one point referred to what is now Turkish Thrace.

1.4.1 Albanian

The oldest dated Albanian documents use the Latin alphabet with Italian-influenced adaptations for sounds not readily represented by a single letter. There is also one early undated document in the Greek alphabet. During the nineteenth century, Arabic, Greek, and Latin alphabets were in competition, each associated with Islam, Orthodoxy, and Catholicism, respectively, although various forms of the Latin alphabet ended up predominating. Cyrillic was also used on occasion (e.g., Pulevski 1875). Nonetheless, the situation was chaotic, with a variety of competing Latin orthographies and all of them competing with Greek and Arabic alphabets. To this can be added the fact that a number of attempts were made at creating a unique alphabet for Albanian (see Elsie 2017). On November 14–22, 1908, an alphabet congress was held in Bitola (Manastir), attended by representatives of all three faiths. The task was to select a single alphabet for use in all Albanian schools and publications as an essential component in the quest for national unity. The delegates were able to agree that Albanian should be written in a Latin alphabet, as this was seen as associated with modernization and not just Catholicism, but they were unable to reach a consensus among the competing Latin orthographies and ended up endorsing two possible choices: one was called *Stambol*, which was similar to an alphabet called *Agimi* and followed the principle of one letter per sound, making use of diacritics and Greek letters (e.g., Greek delta <δ> for the voiced interdental fricative); the other was called "entirely Latin" and came to be referred to as the New Latin alphabet. It was almost identical to an alphabet called *Bashkimi* that made use of diacritics and digraphs (e.g., for the voiced interdental fricative).[91] Children were to be taught both the Stambol and New Latin alphabets in Albanian schools (Skendi 1967: 370–373; see Buda et al. 1972 for documents and details and Sh. Demiraj 2004, Aliu 2005, Lloshi 2008 for additional studies). Eventually the New Latin alphabet became the normal one for all Albanian publications, although the epsilon of the Stambol alphabet (see footnote 91) continued in use a bit longer than the other letters.

1.4.2 Greek

The ancient history of Greek literacy was mentioned in §1.2.3.2. The use of alphabets other than Greek for writing Greek is unusual, although Yavanic was written in Hebrew letters, as was Judeo-Greek of Constantinople in the sixteenth

---

91 The names *Agimi* 'the dawn,' *Bashkimi* 'the unity,' and *Stambol* 'Istanbul' referred to literary societies promoting the respective alphabets. The Bashkimi alphabet used the Italian <gn> for the palatal nasal whereas the New Latin alphabet used <nj>. Bashkimi used <c> for the voiceless dorso-palatal stop, New Latin used <q>. Agimi used the acute accent over <n> and <k>, respectively. The Stambol alphabet used a kind of agma <ŋ> and <q>. For schwa, Stambol used Greek epsilon <ε>, New Latin used e+diaeresis <ë>, Agimi used the inverted *e* used by linguists <ə>, and Bashkimi used <e> for schwa and acute accented <é> for /e/.

century (cf. Hesseling 1897). Arabic was used for writing Greek by Cretan Muslims (Kappler 1998a), and Pulevski 1873 uses Cyrillic for Greek.

1.4.3 Balkan Romance

Although Latin is the traditional alphabet for the Romance languages, Romanian was written in Cyrillic from the earliest documents of the 1500s into the twentieth century. In 1860 the Latin alphabet officially replaced Cyrillic, although Cyrillic continued to be used or at least taught into the twentieth century, and a transitional alphabet using a mixture of Latin and Cyrillic was in use c.1858–1862. The use of Cyrillic reflected the dominance of Church Slavonic in Romanian literary and religious life until the rise of Romanian national consciousness in the nineteenth century. Moldavian was written in Cyrillic 1926–1989, and Article 13 of the 1989 constitution of the Republic of Moldova declared the Latin alphabet official. In the unrecognized state of Transnistria, Cyrillic remains official for Moldovan, although the official language of the Republic of Moldova is now Romanian. Aromanian has been written using Greek letters in regions where Greek is the dominant language (or dominant Christian language), and sometimes still is (in Greece), but all Aromanian schools as well as publications outside of Greece use the Latin alphabet. There is some competition between orthographies that are closer to and further from Romanian (Friedman 2001b). To the extent that Istro-Romanian has been written, it uses the Latin alphabet, in an adaptation of Croatian orthography. Judezmo is written using the Hebrew alphabet, but in the Republic of Turkey, publications in Judezmo use an adapted form of the Turkish Latin alphabet. Nar 1985: 93–287, a collection of nearly 100 Judezmo songs from Thessaloniki, also uses Latin characters. Finally, Cyrillic was also used for Judezmo (cf. Dobreva 2016).

1.4.4 Balkan Slavic

The alphabets of the oldest Slavic documents are Glagolitic and Cyrillic. Glagolitic is generally thought to have been the older of the two and is credited to Cyril and Methodius (see §1.2.3.4). This alphabet was used for some of the Old Church Slavonic documents that have come down to us, and it survived in isolated Croatian monasteries into the twentieth century and is sometimes deployed as a symbol of Slavic identity, especially by Croatians, Czechs, and Slovaks (e.g., the old Czechoslovak twenty-crown note depicted, among other things, the Glagolitic alphabet, and a reproduction of a Glagolitic inscription in the cathedral of Zagreb). In Eastern Orthodox Slavic lands, Glagolitic was supplanted by Cyrillic, an alphabet based on Greek uncial with additional symbols (a couple of which were taken from Hebrew) and attributed to St. Cyril in the Middle Ages but not actually invented by him. Meanwhile the Catholic Slavs used the Latin alphabet. In the nineteenth century, Vuk Karadžić reformed the Serbian Cyrillic alphabet on the principle of one letter per sound,

and this alphabet served as the model for the Macedonian alphabet, which became official on May 3, 1945 (see Friedman 1993b). Bulgarian adapted the more conservative Russian Cyrillic, with a few changes, but did not follow the alphabetical reforms instituted in Russia after the 1917 Revolution until 1945. Prior to codification, Macedonian was written using the Greek alphabet in regions that are now in or near Greece, and Pomak has also been written using the Greek and Latin alphabets (Theocharidis 1996ab; Kokkas 2004; see also §1.2.3.4). Both Serbian and Macedonian had official transliterations in the second Yugoslavia, and still do, although the official transliteration for Macedonian has changed slightly (see User's Guide). The Arabic alphabet (*aljamiado*) was also used by Muslim Slavic speakers, mainly in Bosnia. A variant of Cyrillic known as Bosančica was in use mainly by Catholics and Muslims in Bosnia during the Middle Ages.

1.4.5 Romani

Romani has been written in the Latin, Arabic, Greek, and Cyrillic alphabets, but at present the overwhelming majority of Romani publications use some form of the Latin alphabet. Although an international standard alphabet has been proposed (Cortiade et al. 1991), in practice Romani orthographies tend to follow the orthography of the dominant language of the country in which the publication is produced. EU-sponsored publications tend to follow the Cortiade orthography. In North Macedonia, an alphabet similar to that used for the Latinization of Macedonian Cyrillic is used (see Friedman 1995a). A similar alphabet is also in use in most Bulgarian publications, while publications in Russia still tend to use Cyrillic.

1.4.6 Balkan Turkic

The oldest Turkic alphabet is a form of Runic used for the Orkhon inscriptions (see §1.2.3.6). During the Ottoman period, Turkish was written using the Arabic alphabet, although Turkish-speaking Greek Orthodox Christians used the Greek alphabet, called Καραμανλίδικα, and the Armenian alphabet was used for Turkish-speaking Armenians in Istanbul.[92] During the Ottoman period, Cyrillic was sometimes used in books and manuscripts from Slavic areas (e.g., Pulevski 1875; cf. also Hazai 1963; Kappler 1998a, 1998b), but in general the Arabic alphabet predominated (but cf. Csató et al. 2016). In 1928, Mustafa Kemal Atatürk made the Republic of Turkey switch from the Arabic to the Latin alphabet, a move aimed at secularization and modernization, and this is the alphabet used for Turkish outside the republic as well (Heyd 1954; Lewis 2000). Gagauz was written using Greek, Cyrillic, or Latin letters prior to 1918.

---

92 During the late nineteenth and early twentieth centuries, when the Turkish-language press was significantly censored by the ruling regime, Armenian-alphabet Turkish papers were ignored by the regime, and lessons in the Armenian alphabet were offered to and taken by non-Armenian Turkish-speakers who wanted to read uncensored news (Cankara 2014).

It was written using Cyrillic 1918–1932, Latin 1932–1957, Cyrillic 1957–1996, and is currently written in a Latin orthography based on that of Turkish (see Balta et al. 2018 for additional details).

1.4.7 Georgian and Armenian

Although not a Balkan language, Georgian was used in the monastery of Bačkovo (see §1.2.2.4) in the Rhodope mountains of Bulgaria, where two inscriptions in Georgian (using the Georgian script) can still be seen. The Armenian alphabet has also been used for Armenian, e.g., in Bulgaria, as well as for Turkish in Istanbul, and Armenian is still taught in Greece (Adamou 2008, see §§1.2.2.3–4) and there are Armenian communities in all the Balkan countries.

1.4.8 Orthographies in the Book

As explained in the User's Guide (p. xxxiii), if the language is a nation-state language with a standard Latin orthography or transliteration, then that orthography or transliteration is used here. For OCS, Bulgarian, and East Slavic, the transliteration of Cyrillic in common use in Slavic linguistic publications is used. If the language is a minority language, then the orthography that is official in a country where the language is recognized as such is used. See the User's Guide for discussion of our choices for Greek, Judezmo, and Meglenoromanian as well as decisions concerning dialect material.