2   History of Balkan Linguistics

2.0   Introduction

The field of Balkan linguistics, like all areas of human scholarly endeavor, has developed in two contexts: the external context that defines it as a discipline of study (cf. Gal & Irvine 1995; Irvine & Gal 2000; Gal & Irvine 2019) and the internal context, which is to say the ongoing conversation among its practitioners. In this chapter, we concentrate mainly on the internal history of the discipline, but also give some indication of the context in which it developed.

2.1   Pre-History

The first gleam in the eye of Western Europe that led to the birth of Balkan linguistics is often identified in the section of Thunmann's 1774: 169–366 history of the Eastern European peoples in which he discusses the language and history of the *Albanen* and the *Wlachen*. The term *Albanen* can be unproblematically translated as Albanian without further comment, but Thunmann 1774: 176 himself discusses his choice of the term *Wlachen*, and it is worth noting. He observes that the appellation *Kutsovlah* 'limping Vlah,' used by the Greeks, is a term of abuse, and he labels the form *Wallachen* 'Wallachians' as incorrect (*unrichtig)* despite its being known in Western Europe.[1] Thunmann calls the Balkan Romance speakers south of the Danube *Thracian Vlahs* (*Thracische Wlachen*), and those north of the Danube *Dacian [Vlahs]* (*Dacische*). It was Thunmann (p. 240) who first claimed the Albanians as descendants of Illyrians and "their neighbors the Vlahs" (referring here to Aromanians) as descendants of the Thracians, at the same time rejecting as unacceptable the confusion of Vlahs with Bulgarians or of Albanians with Slavs or people from the Caucasus (p. 170). He also observed that the two peoples were related or intermixed (p. 254). In this section (pp. 181–238), Thunmann also re-published a trilingual Greek–Albanian–Aromanian dictionary (to which he added Latin glosses) by Theodore Kavaliotis of Moschopolis (Aro Moscople, now Alb Voskopoja), a protopriest who originally published the book in 1770 in Venice (see Hetzer 1981).[2] Thunmann noted some of the shared vocabulary between the two languages but was unable to comment on their

---

1 In fact, *Wallachian* is derived via East Slavic in the same way that *Vlah* comes from South Slavic (see §1.2.3.3).

2 It is interesting to note that Thunmann 1774: 178 identifies Moschopolis as being in Macedonia. (The Macedonian form of the name is Moskopole.)

65

grammars (p. 170). It was his identification of Albanian and Balkan Romance with the languages of classically attested peoples and his suggestion that they were related that laid the groundwork for the substratum hypothesis of what became Balkan linguistics, and thus Desnickaja 1970: 46 identifies Thunmann's work as seminal for the field.[3] However, Kostov (1999–2000) and, following him, Asenova (2022: 3) have argued that Cantemir (1716), who briefly describes contact of Aromanian with Greek, and Albanian in terms of "gibberish" (*kauderwälische Sprache*), should be taken as the earliest recognition of Balkan language contact. Be that as it may, we can also mention here Leake 1814: 380:

> It is fair to presume, that the extensive colonization of the Sclavonians in Greece had a proportionate effect upon the vernacular dialects of the country. There is some evidence of this influence in the Albanian and Wallachian, in which the annexing of the article at the end of the nouns, and several other leading features of grammar together with a great similarity of idiom, seem to denote, that from whatever source these languages were originally derived, they were moulded into their present form about the same period, and adapted to the usages of speech of the same great family which had established itself throughout the entire continent of European Turkey. The corruptions which Greek has undergone, may perhaps be chiefly ascribed to the influence of the same great revolution in the population of the South-East of Europe.

This formulation is close to the idea of contact-induced change, and, interestingly enough, attributes it to a Slavic adstratum.[4]

---

3 We can also note here the publication, also in Venice, in 1793 or 1794 and/or 1802, of a quadrilingual dictionary by Daniil Moschopolitis (Alb Voskopoja), which was the first modern work providing parallels using all four of the classic Balkan linguistic groups: In Daniil's terminology: Ῥωμάϊκα, βλάχικα, Ἀλβανίτικα, Βουλγάρικα 'Romaic, Vlah, Albanian, Bulgarian.' (See Kristophson 1974 on problems of editions and dating; see also Leake 1814; Pogorelov 1925; Ninčev 1977; Konstantakopoulou 1988; Stylos 2011; Markovikj 2014.) What Daniil called Romaic is today Modern Greek, while the dialect labeled Bulgarian has been identified as the Macedonian dialect of Ohrid (Koneski 1967: 28). These first comparative lexicons had as their goals the Hellenization of non-Greek-speaking Balkan Christians – see Tsitsipis 1998 on the ideological underpinnings of such efforts – and were quite in contrast to the comparative works of later periods (e.g., whose interest was purely academic, or Pulevski 1875, whose purpose was just the opposite of Daniil's). A portion of Daniil's verse introduction is worth quoting here as illustrative:

> Ἀλβανοὶ, Βλάχοι, Βούλγαροι, Ἀλλόγλωσσοι χαρῆτε,
> Κ᾽ἑτοιμασθῆτε ὅλοι σας Ῥωμαῖοι νὰ γενῆτε.
> Βαρβαρικὴν ἀφήνοντες γλῶσσαν, φωνὴν καὶ ἤθη,
> Ὅπου στοὺς Ἀπογόνους σας νὰ φαίνωνται σὰν μῦθοι.

> 'Albanians, Bulgars, Vlachs and all who now do speak
> An alien tongue rejoice, prepare to make you Greek,
> Change your barbaric tongue, your customs rude forego,
> So that as bygone myths your children may them know.'

<div align="right">(cited and translated in Wace & Thompson 1914: 6)</div>

4 See Schaller 1999 for other publications that hinted at this or that relevant aspect of what would become Balkan linguistics.

2.2   Beginnings

Although the field of Balkan linguistics traces its beginnings to clearly articulated positions on the outer periphery of the Balkan world, i.e., Austria-Hungary and, to some extent, Germany (and later even further afield, to France, England, and Russia as well), eventually the production of knowledge extended into the Balkans itself, with native scholars writing about their own languages and those of their neighbors and/or fellow-countrymen. The beginnings did not give birth to a sudden spurt of growth, however, but rather gestated for a few decades. It was only toward the end of this period that relevant publications began to be produced in significant numbers. Moreover, the fitful growth of Balkan linguistics could not help but be affected by the political upheavals that restricted – but also in some ways enhanced (see §2.2.4) – access to the regions where the languages were spoken.

2.2.1 J. Kopitar

The Slovene linguist and Imperial Austrian censor Jernej[5] Kopitar is generally credited with the first observation pointing clearly in the direction of the development of Balkan linguistics. Kopitar 1829 begins with Thunmann's observations, then surveys the sources available on Albanian (*Albanesische*), Balkan Romance (*Walachische* – north of the Danube and in Macedonia), and Balkan Slavic (what he calls *Bulgarische*). This is followed by a discussion of Romanian Cyrillic and Latin orthographies. He then turns to the question of why *Walachische* differs so much from the other Romance languages and concludes that while the Western European Romance languages were subject to the relative homogeneity of Germanic influence – he claims that they adopted the preposed definite article based on the German (and Greek) model and have more or less German syntax – *Walachische* is basically Vulgar Latin lexicon (*material*) superposed on Thracian grammar (*form*), while Illyrian (Albanian) kept both form and material (Kopitar 1829: 85). The only illustrative item that he cites is the postposed definite article. Observing that postposed articles also occur in Basque and Scandinavian, Kopitar 1829: 86 argues that the similarities between the two "fraternal and neighboring peoples" (*Bruder- und Nachbarvölkern*), i.e., Albanian and Balkan Romance, are in their total grammatical structure. And this *form* was so indestructible (*unvertilgbar*) that it also affected Balkan Slavic, which, under the influence of Balkan Romance, replaced the Slavic case inflections with the postposed definite article. At this point Kopitar 1829: 86 makes his oft-quoted formulation, cited at length in Sandfeld 1930: 11 and briefly in many sources (e.g., Asenova 2002: 7–8; Feuillet 1986: 7; also Friedman 1997a) and worth reproducing here:

> So daß also, noch bis auf diese Stunde, nördlich der Donau in der Bukowina, Moldau und Walachey, Siebenbürgen, Ungern, ferner, jenseits der Donau, in der

---

5 Actually, the form is *Jarnej* in the dialect of his native village of Repnje (as inscribed on a plaque on the house of his birth); his Latin name was *Bartholomeus*.

> eigentlichen Bulgarey, dann in der ganzen Alpenkette des Hämus, in der ausgedehntsten alten Bedeutung dieses Gebirges, von einem Meere zum andern, in den Gebirgen Macedoniens, im Pindus und durch ganz Albanien *nur eine* Sprach*form* herrscht, aber mit dreyerley Sprach*materie* (davon nur eine einheimisch, die zwey andern fremdher, von Ost und West eingebracht sind). [. . .] Also noch sechs Millionen Alt- und Neu- Thracier zwischen den drey Millionen Griechen im Süden und den funfzig Millionen Slaven im Norden.

> And thus, up until today, north of the Danube in Bucovina, Moldavia, Wallachia, Transylvania, Hungary, and beyond, on the other side of the Danube, in Bulgaria proper, and in the entire chain of the Haemus mountains, in the most extended old meaning of these mountains, from one sea to the other, in the mountains of Macedonia, in the Pindus, and in all of Albania, only one grammar dominates but with three lexicons (of which only one is native, the other two having been brought in from outside, from east and west) [. . .] Thus six million old and new Thracians between three million Greeks in the south and 50 million Slavs in the north.

This is followed by some examples, e.g., Latin *homo* and 'Western Romance' *el hom* ("after German *der mensch* with the article preposed") compared with *om ~ omlu* in Macedonia (Aromanian), *omul* in Dacia (Romanian), *tschovek ~ tschoveko* or *tschovekot* in *Bulgarisch* (Balkan Slavic), and Albanian *njerí ~ njeríu*. There follows a discussion of vocabulary and nine Balkan Romance sound changes (rhotacism, various palatalizations and affrications, and changes of velars to labials) compared with Greek, Slavic, and Western Romance. Then, restating his characterization of Albanian, Balkan Romance, and Balkan Slavic (p. 95) as:

> drey lexikalisch verschiedenen, aber grammatisch identischen Sprachen, die vom untersten Donauthale an längs des ganzen Hämusgebirges von Meer zu Meer zwischen den Griechen und Slawen die Grenzscheide

> three lexically different but grammatically identical languages that form a barrier between Greek and Slavic from the lowest Danubian valleys all across the entire Haemus Range, from sea to sea

Kopitar (loc. cit.) gives a translation of the Parable of the Prodigal Son (Luke 15: 11–32) in Serbian, in what he calls *Bulgarisch* ("Bulgarian," but actually the Razlog dialect of Macedonian), in Romanian – the same text in both Cyrillic and Latin transcriptions – what he calls *Macedo-Walachisch*, and in Albanian (Tosk). His purpose in including the Serbian is to make that much clearer the Slavic character of the Serbian and the Balkan (or Balkan Romance) character of the "Bulgarian." After a series of comments on the Albanian lexicon, Kopitar ends by noting that beside the postposed article, the replacement of infinitive by subjunctive and the formation of the future by means of 'want,' which also spread to Greek and Serbian, and the construction Alb [*janë*] *të tutë* = Rmn *ale tale* [*sûnt*] '[they are] yours' ([*sind*] *die deinigen*) are shared grammatical features. Although not an explicit theory of Balkan linguistics, Kopitar's formulation is nonetheless the *terminus a quo* of the field, insofar as it is the first statement to point explicitly to the grammatical commonalties that as such are the key to the concept of areal

linguistics, as opposed to typological or genetic linguistics (cf. Hamp 1977a; J. Greenberg 1957; Jakobson 1958/1962).

2.2.2 A. Schleicher

August Schleicher, the founder of the *Stammbaum* 'stem tree' (i.e., 'family tree') theory of genetic or genealogical linguistic relationships, which treated languages as living organisms and their relationships as biological lines of descent, also commented on the same languages as Kopitar, and like him, placed special emphasis on the postposed definite article (Schleicher 1850: 143):

> Es ist eine bemerkenswerte Erscheinung, dass um die untere Donau und weiter nach Südwesten sich eine Gruppe aneinandergränzender Sprachen zusammengefunden hat, die bei stammhafter Verschiedenheit nur darin übereinstimmen, dass sie die verdorbensten ihrer Familie sind. Diese missrathenen Söhne sind das Walachische in der romanischen, das Bulgarische in der slawischen und das Albanesische in der griechischen Familie. Das Verderbnis zeigt sich in der nördlichsten Sprache, der zuerst genannten, noch in einem geringen Grade, mehr schon in der mittleren, dem Bulgarischen, und hat in der südlichen, der Albanesischen einen ihre Herkunft fast völlig verdunkelnden Grad erreicht. Alle drei stimmen besonders darin überein, dass sie den Artikel an das Ende der Nomina anhangen.

> It is a noteworthy phenomenon that along the lower Danube and further to the southwest, a group of propinquitous languages has coalesced that, being of different lines of descent, agree only in the fact that they are the most corrupt in their families. These ill-bred sons are Wallachian in Romance, Bulgarian in Slavic, and Albanian in the Greek family. The corruption appears in the most northerly, first-named language only to a limited degree, more in the middle one, Bulgarian, and has almost completely obscured the origin of the southernmost, Albanian. All three agree especially in that they attach the article to the end of the noun.

On the basis of this statement, Simpson 1994: 210 credits Schleicher with being the first to recognize the Balkan languages as constituting a sprachbund, but as Kruse 2000/2002: 4.3.2 observes, it is unclear whether or not Schleicher views this "corruption" as "infectious" or mere coincidence. And even if we credit Schleicher with attributing the "degrees" of "corruption" to an unspecified and untheorized substratum, the formulation does not really differ significantly from Kopitar's areal presentation except in its ideology (Bauman & Briggs 2003: 1–18, 97–225).

2.2.3 F. Miklosich and H. Schuchardt

The next significant figure in the history of Balkan linguistics is another Slovene linguist and Slavist, Franz Miklosich, who, among many honors, held the first chair in Slavic Philology at the University of Vienna. Miklosich 1862: 4 refers approvingly to Kopitar's 1829 explanation for the similarities of the Balkan languages and

expands on the number of features he identifies as resulting from the contact-induced restructuring on the basis of a substratum. For the most part, Miklosich's additions were phonological, but in terms of morphosyntax, we give here his list of features (1862: 6–8) and the languages as he cited them with references to the relevant sections of our work:

(1) future with 'want' + infinitive: Serbian, Bulgarian, Modern Greek, Tosk Albanian, whereas Geg has the Romance type with 'have'[6] [§6.2.4.1]
(2) lack of infinitive, with replacement by a finite verb plus a conjunction: Bulgarian, Modern Greek, Albanian, often in Serbian; Romanian has both modes of expression[7] [§7.7.2.1]
(3) merger of genitive and dative: Bulgarian, Romanian, Modern Greek, Albanian [§6.1.1.2]
(4) the un-Romance postposing of the definite article: Bulgarian, Romanian, Albanian [§6.1.2.2.1]
(5) the "prominence" of schwa (Bulgarian, Romanian, Albanian), including reduction of unstressed *a* to schwa [§5.4.1.6]
(6) nasal anlaut (i.e., as syllable onsets before stops in clusters),[8] also loss of *l* before *i*: Romanian, Albanian; also *št* as a common combination: Romanian, Albanian, Bulgarian [§5.4.4.1, §5.4.4.8, §5.4.4.5]
(7) interchange of /n/ and /r/: Romanian, Albanian [§3.2.2.7, §5.4.4.10.5]

Of lesser importance for Miklosich among phonological features, but noted nonetheless, were:

(a) /r/ for /l/: Romanian, Modern Greek, Albanian, and occasionally in Bulgarian [§5.4.4.9.1]
(b) raising of unaccented /o/ to /u/: Bulgarian, Romanian, Albanian [§5.4.1.5]
(c) ea > e when followed by front vowel: Bulgarian, Romanian[9] [§5.4.3.7]

Finally, he added two morphosyntactic phenomena:

(8.1) doubled object pronouns: Bulgarian, Romanian, Modern Greek, Albanian [§7.5.1]

---

6 *Infinitive* here is to be understood only in historical terms.
7 Note that Geg has an infinitive and traces of the Common Slavic infinitive survive in Bulgarian.
8 We indeed intend "syllable" here, not "syllabic"; words such as Albanian *mbret* 'emperor' are monosyllabic so that the *m* is the syllable onset and not a syllabic element on its own (thus, Thomason 2001: 108 misstates this feature).
9 This is the notation that Miklosich uses, based on the Romanian vowels in question (as in *seară*/*seri* 'evening/evenings') although there actually is no /ęa/ as such in Bulgarian (rather /ja/ or /'a/, i.e., /a/ with palatalization of the preceding consonant, depending on one's analytical preferences (in some dialects, the realization of this /a/ is [æ] (which may or may not be phonemic, depending on specifics)). This is Miklosich's representation for the outcome of Common Slavic */æ/ (Glagolitic and Cyrillic *jatь*: <Ⱑ> and <ѣ>, respectively), the source of the /ja/~/e/ alternation in the eastern Bulgarian dialects that served as the basis of the standard language, in which the development was /e/ except under stress when there was a following syllable that had an historical back vowel (thus, Thomason 2001: 108 misstates this feature; see §5.4.3.7).

(8.2) formation of teens with 'ten' and the digits with an intervening preposition: Bulgarian, Romanian, Albanian [§4.3.2.2]

His other observations are connected with the lexicon. It is worth noting here that in contradistinction to earlier work, Miklosich accords more attention to Greek, while mentioning Serbian.

German linguist Hugo Schuchardt introduced the *Wellentheorie* 'wave theory' (Schuchardt 1868: 49; Alvar 1967: 82–85, as cited in Campbell 1998: 189).[10] He added two items to Miklosich's list, namely the change of velar to labial before dental (Latin *luctus* > Romanian *luptă*, Albanian *luftë* 'war') and the Romanian 1SG *am* 'I have' as corresponding to Albanian *kam* 'idem.'[11] Other linguists of this period accepted Miklosich's formulation; see Meillet 1921; Bartoli 1906; Hasdeu 1886; and Weigand 1888.

2.2.4 Data Gathering

From the point of view of Balkan linguistics, the latter part of the nineteenth century and the early twentieth century are characterized by classic works relating to specific Balkan languages or specific aspects of individual or pairs of Balkan languages. A complete list of works would require a separate volume, so we restrict ourselves to a few of the most important authors and works published between the landmarks Miklosich 1862 and Sandfeld 1930 as well as some bibliographies with additional references for this period and beyond. Some significant language-specific works from the period after 1930 are also included here.

For Albanian we can note studies by Hahn 1854; Jokl 1911, 1923; Meyer 1883, 1884, 1892, 1896, 1897; and Miklosich 1870, 1871ab;[12] grammars/chrestomathies by Dozon 1879; Kristoforidhi 1882; Lambertz & Pekmezi 1913; Leotti 1916; Meyer 1888b; Pedersen 1895b; and Pekmezi 1908; dictionaries by Bashkimi 1908; Kristoforidhi 1904; and Meyer 1891. The bibliography by Hamp 1972a is

---

10 J. Schmidt 1872 developed this theory, which challenged the then-prevailing *Stammbaum* ('family tree') model by basically modeling linguistic change as innovations moving outward from a center within a group rather than as divergences from a common ancestor. Although this type of spatial model is appropriate for areal linguistics, Schmidt's object was dialectal differentiation among members of a "genetically" related family. For recent research on Schuchardt, see <http://schuchardt.uni-graz.at/>; see also Gal 2015 on Schuchardt and multilingualism in the Austro-Hungarian context.

11 Regarding *luptă/luftë*, it must be noted that Albanian *drejtë* < Latin *directus* is sometimes taken as evidence that Albanian Latinisms come from two sources, East Balkan Romance and West Balkan Romance, and both represent loanwords rather than a shared Albanoid-Balkan Romance phonological development. For a careful survey and summary of the arguments, see Hamp 1966. What is at issue with a connection between *am* and *kam* would be the final -*m* as a 1SG marker. Rosetti 1968: 155 rejects the possibility of the influence of Albanian on Romanian *am*, and while unwilling to exclude altogether the possibility of a substrate influence explaining both the Albanian and Romanian, he considers such an explanation useless (*inutilă*) in view of the fact that there is ample evidence both from early seventeenth century Romanian as well as other languages (including colloquial French) for the substitution of first plural for first singular.

12 We can include here Tagliavini 1937 for the Geg dialect of Arbanasi (Itl Borgo Erizzo) near Zadar, the only diaspora Geg dialect of Albanian that is well attested.

thorough for the period covered, and Hetzer & Roman 1983: 117–138 provides some additional material. Kastrati 1980 surveys grammars of Albanian 1635–1944, and Ismajli 1982 is a critical edition of the oldest Albanian grammar.

For Balkan Romance, the bibliography in Cazacu 1972 is stronger on post-World War Two work but still useful. Studies, grammars, and descriptions include Capidan 1908, 1925ab, 1928, 1932, 1935; Densusianu 1901; Gamillscheg 1919; Papahagi 1902; Tiktin 1905; Wace & Thompson 1914; Weigand 1888, 1892, 1894–1895; for dictionaries, many of them etymological, there are the works by Cihac 1870, 1879; Dalametra 1906; Tiktin 1903–1925; Paşcu 1925; and Puşcariu 1905 as well as the dialect atlas by Weigand 1909.

For Balkan Slavic, the dialect descriptions by A. Belić 1905; Broch 1903; and Rešetar 1907 are fundamental for Torlak and Stevanović 1935 and Miletić 1940 are contributions to southern Montegrin (Zeta-Lovćen). For Bulgarian we have the historical, analytical, and dialectological work of Conev 1934, 1937, 1940; Meyer 1920; Miletič 1903, 1912; S. Mladenov 1929, 1935; and Todorov 1936, as well as the dictionary of Gerov 1895–1908. We can also mention here Weigand's 1917 grammar as well as Părvev's 1975 detailed survey of the history of Bulgarian grammatical description. For Macedonian dialectology, the fundamental works of this period are A. Belić 1935; Ivanov 1932; Małecki 1934–1936; Mazon 1923, 1936; Mazon & Vaillant 1938; Oblak 1896; and Seliščev 1918, 1929, 1931. The bibliography by Stankiewicz & Worth 1966–1970 is especially useful for this period.

For Greek grammar, the chief works of this period are Chatzidakis 1892; Mirambel 1929; Psicharis 1885, 1929; Thumb 1912; for dialectology we have Dawkins 1940; Dieterich 1908; Heisenberg 1918; Høeg 1925–1926; Thumb 1893; we can also mention here Dawkins 1916, 1937 for Asia Minor Greek and Deffner 1881 for Tsakonian. Meyer 1894 and Murnu 1902 treat loanwords. The relevant bibliographies are by Swanson 1960 and Householder & Nagy 1972.

For Turkish, the grammar by Deny 1921 is the classic pre-reform account of Ottoman Turkish. Redhouse 1890 is the edition that is the basis of what is still the best Turkish–English dictionary: Redhouse 1968; this dictionary has been through numerous reprintings, and supplements have also been published, but for our purposes here this classic edition is still the best. For Turkish lexicon in its Balkan context, Haşdeu 1886; Lokotsch 1927; Meyer 1893; and Miklosich 1884–1890, 1889, 1890 are the major works of the period. The bibliographies by Gülensoy 1981 and Tryjarski 1990 as well as that in Friedman 2003a cover the Balkan Turkish dialects (and see also Johanson 2021: passim).

For the dialects of Romani spoken in the Balkans, Miklosich 1872–1880, 1874–1878; Paspati 1863, 1870; and Pott 1844–1845, are the classic works, and Gilliat-Smith 1915/1916 is an important early dialect classification. Đorđević 1907 can also be mentioned. The bibliography by Bakker & Matras 2003 is an excellent resource.

For Judezmo there are dialect descriptions by Walter 1920; Wagner 1914, 1923, 1925, 1930; Luria 1930; and Crews 1935, and the bibliographies by Studemund

1975 and Sala 1976 as well as Bunis 1975 and the atlas by Quintana Rodríguez 2006.

We can also mention here Horecky 1969, which contains some relevant references to the period, and Schaller 1977, which is really much stronger for the mid-twentieth century than for the time addressed in this section.

Also during this period, the Vienna Academy of Sciences began its series *Schriften der Balkankommission* (from 1900 for the *linguistische Abteilung*), which produced numerous monographs on various Balkan languages and dialects and continues to do so to this day. We can also mention here the *Jahresbericht des Instituts für rumänische Sprache zu Leipzig* (1894–1921), then *Balkan-Arhiv* (1925–1928 and 1976– [new series]). It is not coincidental that a rise in state-sponsored studies of the Balkan languages coincided with Great Power interests in the region. Thus, for example, Wilkinson 1951: 327 notes: "During the War of 1914–1918 [G. Weigand] was commissioned by the German General Staff to enquire into the ethnography of Macedonia. Every facility was granted to him to carry out his task. He was given a staff of six Germans and a number of Bulgarians, but no Serbians served with him for obvious reasons – Serbia was the enemy. His inquiries did not extend into Greek Macedonia as the Germans and Bulgarians were not in occupation of that territory. [. . .] The work was never officially published, but Weigand summarized his ideas in a book published in 1924."[13] See also Axel 2001: 15–21 on the relationship between anthropology and political agendas. Only two works from this time, however, attempted a comparison of all the Balkan languages, both of them doctoral dissertations, Papahagi 1908 and Sandfeld 1900, an excerpt of which appeared as Sandfeld 1902; the former treated parallel phraseology (also Sandfeld 1912), while the excerpt of the latter discussed the replacement of the infinitive. We can also note Michov 1908, which discussed the definite article in Albanian, Romanian, and Bulgarian, and Gilliat-Smith 1915/1916, which noted the commonalities among Albanian, Balkan Armenian, Balkan Romance, Balkan Slavic, Greek, and Romani. This period also saw the start of a number of language-specific journals and dialect series, as well as the publication of folklore collections that served as the basis for Sandfeld 1930 and others.

2.3   Theory and Discipline

Throughout the nineteenth century, and well into the twentieth, the primary tasks of the emergent discipline of linguistics were to gather data and to determine genetic/genealogical relationships.[14] As Martinet observed in his preface to Weinreich 1968: viii: "In spite of the efforts of a few great scholars like Hugo

---

13 It is worth noting that while German scholarship of this period stressed the links of Macedonian with Bulgarian, French scholarship (e.g., Vaillant 1938) stressed the relative distinctness of Macedonian vis-à-vis both Serbian and Bulgarian.

14 The changing perspectives are illustrated by the titles of Sandfeld's seminal works. The 1926 Danish original was 'Balkan philology' but the 1930 French translation was 'Balkan linguistics.'

Schuchardt, linguistic research has so far favored the study of divergence at the expense of convergence." Nonetheless, it was during the 1920s that new approaches to language classification emerged and that works providing the theoretical underpinnings of Balkan linguistics as well as the syntheses of data that led to the construction of Balkan linguistics as a discipline were first published.

2.3.1 N. Trubetzkoy and R. Jakobson

Trubetzkoy 1923 contains the first theoretical formulation of the concept of a *sprachbund* (French *union linguistique*, Russian *jazykovoj sojuz*, English *linguistic league*). Owing to both the obscurity of the original publication and the lesser frequency of knowledge of Russian in Western Europe, Trubetzkoy's 1930 restatement at the First International Congress of Linguists is better known and more often cited. In his original article, having discussed the classical Stammbaum presentation of languages as families, branches, languages, and dialects (and, we can add, having observed that in the case of transitional dialects the means of linguistic science alone are often unable to resolve a quarrel over which language can claim the given dialect), Trubetzkoy 1923 remarks that languages in a given geographic and cultural-historical region can form a group whose resemblances are not due to common ancestry but rather prolonged contact and parallel (i.e., convergent) development. He suggests the term *jazykovoj sojuz* 'language union' for such groups, and, in a footnote, adduces the Balkan languages – which he specifies as Bulgarian, Romanian, Albanian, and Modern Greek – as exemplary owing to the common traits in their grammatical structures. He then goes on to suggest that such unions exist not only at the level of language, but also at the level of family, thus anticipating Jakobson's discussion (see Jakobson 1931a/1962, 1931b/1962, 1938/1962) of phonological affinities, but also such problems as Altaic (see Johanson 2006; cf. also Masica 2001; Nichols 1992). Trubetzkoy 1930 was published as proposition 16 in answer to general theme II: "*Etablissement et délimitation des termes techniques. Quelle est la traduction exacte des termes techniques dans les différents langages (français, anglais, allemand)*?" ('Establishment and delimitation of technical terms. What is the precise translation of technical terms in different languages (French, English, German)?'). The formulation is of sufficient importance to be cited here in its entirety (italics as in the original):

> Viele Missverständnisse und Fehler entstehen dadurch, dass die Sprachforscher die Ausdrücke "Sprachgruppe" und "Sprachfamilie" ohne genügende Vorsicht und in zu wenig bestimmter Bedeutung gebrauchen. Ich schlage folgende Terminologie vor:
> *Jede Gesamtheit von Sprachen, die miteinander durch eine erhebliche Zahl von systematischen Übereinstimmungen verbunden sind, nennen wir Sprachgruppe.*
> Unter den Sprachgruppen sind zwei Typen zu unterscheiden:
> Gruppen, bestehend aus Sprachen, die eine grosse Ähnlichkeit in syntaktischer Hinsicht, eine Ähnlichkeit in den Grundsätzen des morphologischen Baus

aufweisen, und eine grosse Anzahl gemeinsamer Kulturwörter bieten, manchmal auch äussere Ähnlichkeit im Bestande der Lautsysteme, – dabei aber keine systematischen Lautentsprechungen, keine Übereinstimmungen in der lautlichen Gestalt der morphologischen Elemente und keine gemeinsamen Elementarwörter besitzen, – *solche Sprachgruppen nennen wir Sprachbünde.*

Gruppen, bestehend aus Sprachen, die eine beträchtliche Anzahl von gemeinsamen Elementarwörtern besitzen, Übereinstimmungen im lautlichen Ausdruck morphologischer Kategorien aufweisen und, vor allem, konstante Lautentsprechnugen bieten, – *solche Sprachgruppen nenen wir Sprachfamilien.*

So gehört z.B. das Bulgarische einerseits zur slawischen Sprachfamilie (zusammen mit dem Serbokroatische, Polnischen, Russischen u.s.w., andererseits zum balkanischen Sprachbund (zusammen mit dem Neugriechischen, Albanesischen und Rumänischen).

Diese Benennungen, bzw. diese Begriffe sind streng auseinanderzuhalten. Bei der Feststellung der Zugehörigkeit einer Sprache zu einer gewissen Sprachgruppe muss der Sprachforscher genau und deutlich angeben, ob er diese Sprachgruppe für einen Sprachbund oder für eine Sprachfamilie hält. Dadurch werden viele voreilige und unvorsichtige Äusserungen vermeiden.

Many misunderstandings and mistakes arise when the linguist uses the terms "language group" and "language family" without the necessary care and with inadequate definition. I propose the following terminology:

We will call a "language group" every community of languages that is connected by a considerable number of systematic correspondences.

There are two types of language groups to be distinguished:

Groups comprising languages that display a great similarity with respect to syntax, that show a similarity in the principles of morphological structure, and that offer a large number of common culture words, and often also other similarities in the structure of the sound system, but at the same time have no regular sound correspondences, no agreement in the phonological form of morphological elements, and no common basic vocabulary – such language groups we call *sprachbunds* [language unions].

Groups consisting of languages that possess a considerable amount of common basic vocabulary, that show correspondences in the phonological expression of morphological categories, and, above all, display regular sound correspondences – such language groups we call *language families.*

Thus, for example, Bulgarian belongs on the one hand to the Slavic language family (together with Serbo-Croatian, Polish, Russian, etc., and on the other hand to the Balkan sprachbund (together with Modern Greek, Albanian, and Romanian).

These terms or concepts should be strictly distinguished from one another. In establishing the belonging of a language to a given language group, the linguist must clearly and precisely specify whether he considers this language group as a sprachbund or as a language family. Thus will many rash and careless statements be avoided.

Again, it is worth noting here both the inclusion of Modern Greek in the sprachbund and the exclusion of Serbo-Croatian. These are questions already implicit in Kopitar 1829 and ones that will arise repeatedly. Also worth noting are Trubetzkoy's specification of "historical-cultural" in addition to "geographic" in 1923 and *Kulturwörter* in 1930. Although Masica 2001: 239 cautions against

confusing "recent political configurations" with linguistic areas, nonetheless, in the case of the Balkans (and elsewhere) it is precisely such social factors as the political that create the conditions for convergence. Trubetzkoy's formulation of "belonging" to a sprachbund in the same sense as "belonging" to a linguistic family will prove to be problematic (see §2.5.3 and especially Weinreich 1958), but even from the earliest days, the positions of Serbian (or Serbo-Croatian) and Greek were labile (cf. Masica 2001: 241, cited in Chapter 1, footnote 11).

Jakobson's discussion of phonological sprachbunds and the Eurasian sprachbund (see Jakobson 1931a/1962, 1931b/1962, 1938/1962) concentrates on consonantal timbre (basically palatalization including some correlations with front/back vowel harmony), prosody (presence vs. absence of pitch accent or tone), and, in a footnote to Jakobson 1931b/1962: 191, nominal declension. He sets up Eurasia as the center in terms of all these. For nominal declension, Germano-Romance Europe and South and Southeast Asia are the peripheries, in terms of phonological tone, the Baltic and Pacific areas are the peripheries (with West South Slavic [most of Serbo-Croatian and Slovene] as a relic island), while for palatalization the core is roughly the boundaries of the Russian Empire, with the inclusion of eastern Bulgaria.[15] He even goes so far as to suggest that palatalization in Great Russian [*sic*] finds its most complete expression, and it is thus no coincidence that Great Russian is the basis of the Russian literary language, i.e., the language with a pan-Eurasian cultural mission (Jakobson 1931b/1962: 191).[16] It is worth noting, however, that in his formulation of the definition of sprachbund, in which he cites Trubetzkoy's, Jakobson 1931b/1962: 145 also makes it clear that a sprachbund consists of "*dvuh ili neskol'kih*" ('two or several') languages in contact. The point is that the convergent processes that define a sprachbund are of the same type regardless of whether two or more than two languages are affected.[17]

We return to the problems of phonology and sprachbunds in Chapters 3 and 5.

---

15 There is a partial correlation between palatalization and late nineteenth-century Russian territorial aspirations to a "Trans-Danubian Province" in the Balkans, which would have included all or most of modern-day Bulgaria. See footnote 16.

16 In view of what we say in §2.5.1, let us observe here that Jakobson 1931b/1962: 167–168 includes Russian Romani (Xaladytka) among the languages of his Eurasian sprachbund. While Jakobson located Russia at the center of this Eurasian sprachbund, Haspelmath 1998 posits a "Standard Average European" with French, German, Dutch, and North Italian as its center and the rest of Europe as the periphery. Moreover, just as Jakobson's formulation coincided fairly closely with Russia's perceived geopolitical sphere of interest, so, too, Haspelmath's version of the development of a European sprachbund coincides with EU relations of core and periphery. This is not to say that either linguist was attempting to act as a tool of foreign policy (although Jakobson's advocacy of a Russian cultural mission could be read that way), but at the same time, once such works are published they can be adopted and adapted by those with policy goals.

17 This point is also made in Thomason & Kaufman 1988: 95 but bears repeating since recently the idea that a sprachbund cannot be dyadic has been raised without justification. See §3.4 for some discussion.

2.3.2 A. Seliščev and Kr. Sandfeld

We turn now to two key figures in specifically Balkan studies in the first quarter of the twentieth century, the Russian linguist Afanasiy Seliščev and the Danish linguist Kristian Sandfeld.

2.3.2.1 Seliščev

Seliščev 1925 begins with the arrival of the Slavs in the Balkans during the sixth and seventh centuries and a description of the Jireček line. Seliščev then briefly notes a few historical events up to the Turkish conquest, framing them as a struggle between Byzantium and Bulgaria which prevented the isolation of the various ethnic groups by subjecting them to the same fate and necessitating their cooperation, and, therefore, linguistic contact. He also notes influences emanating from the Roman west as well as the Hellenic east, with Romance being particularly strong in Albania and western Macedonia. Then comes the Turkish conquest of most of the Balkans (especially the part of interest to us; see §1.1) in the fourteenth century. He comments on the prolonged and intense influence of Turkish, and his first concrete example consists of the Turkish derivational affixes *-ci (dži)*, *-lik*, and *-li* (with allowances for vowel harmony and adaptation) used for forming agentive, abstract, and attributive nominals, respectively, and borrowed productively into all the Balkan languages (see §4.2.2.3). He then comments on the same phenomenon with respect to the borrowing of the Slavic derivational suffix used to mark female gender, *-ica*, in all the non-Slavic Balkan languages (including, we can add, Balkan Turkish, cf. §4.3.8), and already attested in Greek in the early twelfth century. Likewise, the -σ- of the Greek aorist is used productively in forming verbs in all the non-Hellenic Balkan languages. He then cites two semantic parallels (the same word for 'century,' 'life,' and 'world' and the expression 'sing' for 'read') and refers the reader to Papahagi 1908. We discuss the lexicon in some detail in Chapter 4.

Seliščev then turns to morphology and syntax and adduces the following features:

(1) resumptive clitic pronouns to mark dative and accusative objects. He notes that the phenomenon is regular in the dialects of Macedonia and more rare in eastern Bulgaria

(2) reduplication to indicate quantity, intensity, distributivity, e.g., Alb *bol bol* 'a huge amount,' *pika pika* 'drop by drop' (noting that this also occurs in South Italian and Armenian as well as Sanskrit, where it is known as *āmreḍita* 'repetition,' e.g., *gr̥hé-gr̥hé* 'house by house')

(3) a single word for 'where' and 'whither'

(4) frequent use of the subjunctive

(5) replacement of the infinitive with a personal form of the verb and a conjunction[18]

---

18 It would seem that he is distinguishing between subordinate and independent uses of the DMS (see §7.7.2.1.3.1); he also uses the term 'particle' (p. 50).

(6) use of 'for' + subjunctive to indicate goal, intention, desire
(7) replacement of future with a construction using the present of the verb 'want' and the infinitive and replacement of the infinitive with the subjunctive
(8) postposed definite article (Albanian, Balkan Romance, and Balkan Slavic)
(9) postposed dative–genitive pronominal clitics used to indicate possession
(10) loss of case forms, generalization of the accusative with prepositions and other expressions, merger of genitive and dative, and many other phenomena for Balkan Slavic and Balkan Romance

This is followed by a discussion of phonetic traits, but he notes (p. 49) that these may be independent parallel developments. He then cites several local level phonological phenomena that can be reliably attributed to contact:

(11) velar /ł/ in Frasheriote and Tirana Aromanian, as in Albanian
(12) similar conditions for the change of *î* to *ă* in Meglenoromanian and the neighboring Macedonian dialects
(13) presence of /šč/ (e.g., *ščiu* 'I know') in the Aromanian of southwestern Macedonia under Slavic influence
(14) loss of unaccented initial vowels in the Macedonian dialects east of Thessaloniki, a feature of many Greek dialects

He then lists a few other phenomena from Seliščev 1918, e.g., mellow palatal stops, loss of /x/, loss of intervocalic voiced consonants (especially, *g, d, w/v*) after an accented syllable, which he identifies as Macedonian without indicating the languages or dialects with which they are shared (but see §5.4.4.10.2). Seliščev then notes that the state of our knowledge of the history of Albanian and the development of Romanian is too impoverished to answer the *grandes questions* ('big questions') of origins and influences. He observes that Sandfeld's 1902[19] (and Meyer's and Pedersen's) attribution of the loss of the infinitive to Greek influence, while plausible, is not demonstrable in the absence of equally old texts for the precursors of Albanian and Romanian. With regard to emphatic reduplication, Seliščev observes that the phenomenon is attested as early as the fifth century BCE in Greek, but that this does not permit us to assume that Greek was the first to use the construction, since we know nothing of the ancestor of Albanian, and in modern Albanian such reduplication is more frequent.[20] He then observes that based on the available documentation, Slavic appears to have been the recipient of innovations, except in the case of vocabulary, including derivation (cf. *-ica* noted above). He closes with a discussion of intensive reduplication in Balkan Slavic, arguing that while the textual evidence shows that distributive reduplication with numerals and nouns governed by a preposition could have a Greek model, the intensive-emphatic reduplication of adjective and bare nouns resulted from Turkish influence. This article represents the most thorough treatment of the Balkan

---

19 Actually, Sandfeld-Jensen.
20 This is followed by speculation on Armenian, based on the assumption that it is descended from Phrygian (see §1.2.1.6).

languages at the time and addressed an increasing number of important issues. To the best of our knowledge, it also introduced the term *Balkanism* for those traits common to the Balkan languages as a result of language contact.

2.3.2.2 Sandfeld

A year after Seliščev's article appeared, in 1926, Sandfeld published his epoch-making work in Danish. The epoch did not really begin, however, until 1930 when the French edition was published, thereby making the book available to a wider audience.[21] Sandfeld's was the first treatment to collect and synthesize the many individual studies and textual materials, most of which had been published during the preceding sixty or seventy years, and it is justly credited with establishing the discipline of Balkan linguistics as such. Nonetheless, despite the emphasis on structural similarities from Kopitar to today, Sandfeld concentrates on shared vocabulary, devoting almost half the work (pp. 16–99) to common Balkan loan-words from Greek, Latin, Romance, Albanian, Slavic, Turkish, and Others (words of uncertain etymology possibly of substrate origin, especially Thracian, most of them pastoral, but also Germanic, Hungarian, Celtic, Proto-Bulgar, etc.). His next chapter (pp. 100–162) treats correspondences "outside the lexicon" in terms of pairs of languages, in each case the first generally being imputed as the source of influence: Greek with Aromanian, Albanian, and Balkan Slavic; Albanian with Aromanian and Balkan Slavic; Albanian with Romanian; Balkan Slavic with Balkan Romance and Albanian; and Turkish with all of the above.[22] In most of these sections, he discusses issues of phonology, morphology, syntax, and phrase-ology. The majority of these phenomena, including all those that are phonological and inflectional, are limited to specific areas of contact within the Balkans rather than larger regions. Some are more broadly attested and of a causality different from what Sandfeld was aware of, e.g., the loss of gender distinctions in third singular clitic pronouns in southwestern Macedonian south and west of Struga-Ohrid-Resen-Bitola-Gumendže (Grk Gouménissa)-Enidže Vardar/Pazar (Grk Gianitsa) as far as Kostur (Grk Kastoria), but not Korça, as well as Upper Polog, and part of Poreče (Koneski et al. 1968: 521, 530), reflects the situation not only in Albanian (Sandfeld 1930: 120), but also in Aromanian (Koneski 1981: 133), which is in fact the evident source of the phenomenon in southwestern Macedonia, although Albanian was probably the source in Polog and Poreče. Similarly, some phenomena are represented in all the language groups but not all the languages, e.g., sentence-initial clitic pronouns, which occur in Albanian, Greek, Balkan

---

21 The 1930 French version is, in Sandfeld's own words, not just a translation but a "*deuxième edition, un peu remaniée et notablement augmentée*" ('second edition, a bit reworked and notably augmented'). See Skytte 1994 for an interesting biography of Sandfeld, with a complete bibliography of his works.

22 Sandfeld usually differentiates between Aromanian and Romanian, and also gives separate data for Meglenoromanian in sections dealing with Aromanian, although he also uses *Romanian* as a cover term for Balkan Romance. In general he also distinguishes "bulgare" from "macédo-bulgare" although not with complete consistency, and he uses *Bulgarian* as the cover term for Balkan Slavic. Only rarely does he specify a difference between Geg and Tosk Albanian, and most of the actual data cited are Tosk.

Romance, and Balkan Slavic as a larger category, but not in Bulgarian (or, for the most part, Torlak BCMS, or eastern Macedonian) within Balkan Slavic. Likewise, the so-called Balkan Conditional in its furthest development, i.e., uninflected future particle plus imperfect, is characteristic of Greek, SDBR, Tosk Albanian, and Macedonian (Sandfeld 1930: 105; see now Belyavski-Frank 2003).[23] In the section on Turkish, Sandfeld cursorily cites a few calques and argues that since Turkish is so structurally different from the Balkan languages and moreover such a recent arrival, its influence does not extend beyond the lexical and phraseological. And even in this, Sandfeld 1930: 161 is at pains to argue against Turkish as the source for such commonalties as doubling in an augmentative function.

In his fourth and final chapter, Sandfeld 1930: 163–216 discusses what he calls "general correspondences outside the lexicon." While acknowledging that some items in his previous chapter are widespread, the items of chapter four are those that Sandfeld identifies as specifically "Balkan" and constitutive of "*l'air d'unité*" ('sense of unity') found among the Balkan languages, which he defines as Albanian, "Bulgarian," "Romanian," and Greek, and often also Serbo-Croatian. He excludes Turkish. He begins this section by reviewing some of the sillier proposals for explaining Balkan linguistic convergence (e.g., M. Gaster's 1888 *Turanian*, i.e., Proto-Bulgar, hypothesis), and, citing H. Pedersen and G. Meyer with approval, indicates that he will argue against the substratum hypothesis and in favor of Greek as the source of Balkan linguistic convergences/unity, owing to the superior cultural prestige of the language. The specifics of the argumentation are taken up throughout the present work, and in Chapter 8, and we discuss an expanded inventory of morphological and morphosyntactic Balkanisms that takes Sandfeld's third and fourth chapters together into account in our Chapter 6. For our purposes here, it suffices to list those features that Sandfeld identified as salient Balkanisms in his fourth chapter as a kind of culmination point for the lists that have gone before:[24]

(1) postposed definite article
(2) loss of the infinitive
(3) future formed with 'want'
(4) merger of genitive and dative plus use of dative pronouns as possessives
(5) same form for 'where' and 'whither' plus use of proleptic personal pronouns [including object reduplication]

---

23 Sandfeld only cites future marker + verbal *l*-form for Macedonian, but in fact the construction occurs with the imperfect and 'have' perfect as well.

24 Sandfeld 1930: 171 mentions diphthongization of older **ě* in eastern Bulgarian and *e* in Romanian in stressed syllables followed by a back vowel, but not in those followed by a front vowel, as possibly related, but argues that the reduction of /a/ to schwa in Bulgarian is probably unconnected with the earlier manifestation of the same phenomenon in Albanian and Romanian. However, he observes that as a contact feature, the raising of unaccented /e/ to /i/ and /o/ to /u/ in dialectal Romanian, Aromanian, (eastern) Bulgarian, and northern Greek is "somewhat less doubtful" (*quelque peu douteux*). He also mentions here, and rejects, the notion that the substratum is responsible for the simplification of declension, and he observes that it is necessary to demonstrate that what is "normal" in the development of one language (presumably Western European) can be due to foreign influence in another (Sandfeld 1930: 171–173), i.e., the now-old question of typology (universal tendencies) versus areality (contact-induced phenomena).

(6)   nominal complement treated as accusative object

(7)   negative clause followed by 'and' plus main clause

(8)   other paratactic constructions

(9)   two direct objects (especially with 'learn' and 'ask')

(10) be ten years [old] = have ten years [of age]

(11) 'how' as 'approximately'

(12) phraseology, e.g., the 451 entries in Papahagi 1908

(13) features shared with other [non-Balkan] languages: phraseology, e.g., Italian *senz' altro* = Bulgarian *bez drugo*, Albanian *pa tjetër*, Romanian *fără de alta*, Greek χωρίς άλλο 'without a doubt' (lit., 'without other').

Sandfeld concludes by arguing that it was the Byzantine Empire that constituted the political frame for the Balkan linguistic league, and that Greek was the source for most of the salient features. While subsequent investigation has shown Sandfeld's approach to causality to have been oversimplified, his actual collection, organization, and synthesis of data remain the high point of its kind in Balkan linguistics until the later part of the modern period.

2.3.3 Between *Linguistique balkanique* and *Balkansko ezikoznanie*

From the epochal publication of Sandfeld 1930 until the late 1950s and early 1960s, the study of Balkan linguistics went through a period of slow growth. This is not to say that there was no scholarly activity. P. Skok and M. Budimir's interdisciplinary but short-lived *Révue internationale des études balkaniques* (Belgrade 1934–1938), which set as the goal of Balkan linguistics not only the identification of influences of individual Balkan languages but also the establishment of those Balkanisms characteristic of the sprachbund (Skok & Budimir 1935: 15), published a number of important articles (e.g., Skok 1935; Anagnostopoulos 1935; A. Belić 1936). Collections such as the festschrifts for A. Teodorov-Balan (Dimitrov 1955) and I. Iordan (Cazacu 1958) also contained seminal works by important scholars such as V. Georgiev 1955, A. Graur 1955, M. Caragiu-Marioțeanu 1958, and scholars such as Małecki (e.g., 1935) and Gołąb (e.g., 1956), as well as many others produced important work both before and, especially, after World War Two. (See Asenova 1979 and Schaller 1999 for a more detailed review of the field and Schaller 1977 for additional bibliography.) It was during this period that Latin joined Greek as the sought-after source of Balkan convergence phenomena (e.g., Gołąb 1956).

Not long before Gołąb's 1956 identification of the isogrammatism (see §6.2.3.2.1), Uriel Weinreich in 1953 published the book that grew out of his master's thesis and doctoral dissertation and became a seminal work: *Languages in Contact* (second edition, 1968). Curiously enough, however, the Balkans are almost completely ignored in Weinreich's classic. Aside from passing mentions of Romanian codeswitching and loanwords (Weinreich 1968: 74, 82) – neither of them in a Balkan context – the citation of Capidan's 1925b: 159 account of Meglenoromanian borrowing its first and second person singular markers from

Macedonian (Weinreich 1968: 31–32; but see §2.5.3 below),[25] and a garbled account of borrowing from Greek into Turkish,[26] the only passage connected with the Balkans is the following, at the end of the study (Weinreich 1968: 113):

> Some parts of the world have traditionally formed linguistic whirlpools, and some languages have been exposed more than others to linguistic cross currents.
>   One famous area of multiple language contacts has been the Balkan peninsula. It has fascinated students of interference for decades; since Schuchardt (496),[27] it has served as a storehouse of standard examples for practically every type of interference. Special periodicals devoted to Balkanology have explored the problems of common linguistic and cultural features of this area, and numerous separate studies have been published on the subject.

A footnote refers to Skok and Budimir's journal, Sandfeld's "celebrated works," and a few other items. Although Weinreich's reference to the Balkans occurred in a closing section entitled *Multiple Language Contacts as a Favorable Field of Study*, Balkan linguistics continued to develop as a discipline referred to but not investigated in what became the field of contact linguistics, a field for which Weinreich 1953 is in a position in some ways comparable to that of Miklosich 1862 for Balkan linguistics.

It was also during the 1950s that Balkan national linguistic journals such as *Makedonski jazik* ([North] Macedonia, 1950–), *Izvestija na Instituta za bălgarski ezik* (Bulgaria, 1951–), *Studime filologjike* (Albania, 1957/1963–), *Studii şi Cercetări lingvistice* (Romania, 1950–), and others, began publication. While these journals are not specifically aimed at problems of Balkan linguistics, many prominent Balkanists have published in them, and a number of their articles either have addressed such problems or supplied information for those who have done so. Similarly, university and academy yearbooks, journals, and series, e.g., the series of seven B.A. theses (*Diplomni raboti*) published by the University of Skopje, published works of direct or indirect relevance to Balkan linguistics.

Nonetheless, compared with the explosive growth from the 1960s onward, the decades from 1930–1959 are marked by a much lower rate of productivity. If we take a representative sample of the literature in or relevant to Balkan linguistics between 1930 and 1960, e.g., the bibliography in Asenova 2002: 334–370, we note that the decade 1950–1959 produced about twice as many works as that of 1930–1939, and moreover three-quarters of the 1950s' production occurred during the second half of the decade.[28] Furthermore, all those works taken together account for about a tenth of the works cited, the overwhelming majority having been produced during the subsequent decades.

---

25 Capidan's (and thus Weinreich's) example is actually quite problematic; see §6.2.1.1.3.
26 At issue is not "[t]he Greek agentive suffix *-ci*, transferred into Turkish" but rather a shape of the Turkish agentive suffix as invariant *-cis* in Istanbul market slang under the influence of the Greek -τζης, itself from Turkish *–ci/-cü/-cı/-cu* (although these would all be neutralized to *-dži* in West Rumelian Turkish); see Spitzer 1936.
27 The item referred to is Schuchardt 1884.
28 See also Schaller 1977. For obvious reasons, Asenova's 2002 bibliography contains only eight works from the 1940s. Of these, four are on Greek grammar and dialectology and one each on Albanian, Bulgarian, Latin, and Romanian.

2.4 The Modern Period

The significant increase in the production of works on Balkan linguistics during the latter half of the 1950s was augmented by the publication of a number of journals devoted to the Balkans, and in whole or in part to Balkan linguistics. The first of these was *Balkansko ezikoznanie* (Sofia, 1959; also known as *Linguistique balkanique*) published by the Bulgarian Academy of Sciences. Other such journals have been *Balkan Studies* (Thessaloniki, 1960–), *Zeitschrift für Balkanologie* (Wiesbaden, 1963–), *Godišnjak – Centar za balkanološka ispitivanja -Akademija nauka i umjetnosti Bosne i Hercegovine* (Sarajevo, 1963–1989, 1997–), *Révue des études sud-est européennes* (Bucharest, 1963–), *Études balkaniques tchécoslovaques* (Prague 1966–1972), *Balcanica* (Belgrade 1970–), *Balkanistica* (Columbus [OH], Bloomington [IN], Oxford [MS] 1976–), *Balcanica posnaniensia* (Poznań, 1984–), and others (see Asenova 1979).

The upsurge in Balkan linguistic studies that began in the late 1950s and increased dramatically throughout the rest of the twentieth century has continued unabated. A milestone in the affirmation of Balkan linguistics as a discipline in its own right was the founding in Bucharest in 1963 of the International Association for Southeast European Studies (known by its French acronym: AIESEE). Although overwhelmingly dominated by historians, and, well into the 1990s, by a social science orientation typical of the Soviet Union and its bloc, AIESEE has nonetheless contributed to the conversation among Balkan linguists both in print and in its congresses in Sofia (1966), Athens (1970), Bucharest (1974), Ankara (1979), Belgrade (1984), Sofia (1989), Thessaloniki (1994), Bucharest (1999), Tirana (2004), Paris (2009), Sofia (2014), and Bucharest (2019). Another congress is planned, as of this writing, for Skopje (2025). A major result of the research from the years preceding it was volume VI of the proceedings of the first congress of AIESEE (Gălăbov et al. 1968). This volume stands as a monument of sorts to the state of Balkan linguistics of that period, with virtually every scholar in the field at the time represented in it (79 articles and 108 commentaries by over 100 scholars).

The last third of the twentieth century, dating from the publication of volume VI of the proceedings of the first AIESEE Congress, was a period of tremendous productivity for scholarship on Balkan linguistics that has continued into the twenty-first. The number of relevant books and articles, e.g., Joseph 1983a on the loss of the infinitive, Civ'jan 1979 on syntax, and Sawicka 1997 on phonology, to take just three examples of monographs of three different types in each of three decades (see also footnote 1 in the Introduction), surpasses the total production of the preceding hundred years. Publications connected with the *Malyj dialektologičeskij atlas balkanskih jazykov* (in addition to references in the Introduction, see Sobolev 2001b; Ylli & Sobolev 2002, 2003) promise to add a significant quantity of comparable syntactic and lexical data. Also worthy of note is the fact that while in the past Balkan linguistics has been discussed at general linguistic and phonetic conferences (e.g., Trubetzkoy 1930; A. Belić 1936), the International Congress of Slavists has done more, and has a standing committee on Balkan linguistics founded at the eleventh Congress, in

Bratislava in 1993, which has continued to be active since then (but cf. also the dictionary already proposed in Batowski 1939).

Schaller's 1975 handbook was the first attempt to take stock of the field in almost half a century, and as such generated considerable attention and no less than thirteen reviews and review articles (see Joseph 1987a). As noted in the Introduction, it was followed by eleven other handbooks of various types and orientations, from brief syntheses to voluminous analyses to compendia of articles on individual topics, and we discuss the merits and views of these studies as appropriate in the chapters which follow.[29] Earlier in this chapter we adduced the chief Balkanisms identified from the beginning to the end of what can be called the pre-modern period of Balkan linguistics as represented in the works of Kopitar, Miklosich (plus Schuchardt), Seliščev, and Sandfeld. While keeping in mind Jakobson's 1958/1962: 524 admonition that system and not mere inventory must be the basis of our study, it is nonetheless convenient to use lists as a kind of shorthand for the systemic relations that can yield the most insights.[30] By way of comparison with the preceding periods, we therefore close this section with a slightly modified version of the tabular summary given in Asenova 2002: 294–295, which is the best of the recent summaries of its kind.[31] We address all the nuances of these phenomena, and many others, in the chapters that follow.[32]

29 The following brief observations can be made here. Feuillet 1986 is intended as a brief college-level introduction rather than as a thorough scholarly treatment. Steinke & Vraciu 1999 was written in 1974, contains little actual data, and ignores English-language scholarship. Schaller 1975 is flawed by errors of methodology and fact, as are Banfi 1985 and Tomić 2006 (see Introduction, footnotes 3 and 14), which latter only treats morphosyntax. Solta 1980 is weakened by adherence to the substrate hypothesis, which is founded on assumptions that are unsupported by any concrete evidence (see §1.2.1), while Reiter 1994 is excessively mechanistic and short on actual data. Haarman 1978 is not a handbook but rather a study of Latinate vocabulary, as is Banfi 1991, which focuses on the Romance element during the medieval period – an important topic, to be sure, but only one topic. Sh. Demiraj 2004 is in many respects a condensation of Sh. Demiraj 1985 with non-Albanian data added, some of it corrected from the first (1994) edition. Desnickaja & Tolstoj 1990, 1998 fail to synthesize the descriptions of individual languages. Hinrichs 1999a is not a synthesis but rather an encyclopedic compendium of articles covering in significant detail almost all the traditional aspects of Balkan linguistics as well as questions of onomastics, history, and culture. Less traditional topics such as contact-induced change in the Balkan dialects of Turkish and of Jewish languages are absent. Feuillet 2012 treats Balkan Slavic, Balkan Romance, Albanian, and Greek from a viewpoint of synchronic comparative grammar; it is based mostly on standard languages, and discusses both Balkan and non-Balkan, i.e., language-specific, features.

30 What we do not want to do is to fetishize the labels for these systemic manifestations, assigning numeric values to them and tallying up the number of points a language "scores" (Friedman 2000b as misinterpreted by Lindstedt 1998, 2000; cf. also van der Auwera 1998); a critique of this approach is given in §3.3. Rather we see these labels as standing for complex interrelations that must be elucidated in their overall contexts.

31 The first edition of Asenova 2002 included a section of seven full and five partial *isosemic* correspondences, i.e., calqued or parallel semantic shifts or equations on p. 232, e.g., 'heart' > 'courage' and 'written' > 'fated.' These were omitted from the second edition. We have also altered her *Bulgarian* and *Romanian* to *Balkan Slavic* (BSl) and *Balkan Romance* (BR), respectively. While Asenova 2002 is marred by Bulgarian and Romanian nationalism vis-à-vis Macedonian and Aromanian and is also mostly restricted to standard languages, it is the first such handbook to contain detailed historical analyses in addition to the comparative material, and we have occasion to refer to it frequently.

32 We have added the source languages for the full Balkanisms and the languages lacking a given feature for the partial Balkanisms.

Full Balkanisms	Partial Balkanisms
	Sound System
reduction of unstressed vowels	stressed schwa [-Grk, -WMac, -GegAlb]
	mp/*nt*/*nk* > *mb*/*nd*/*ŋg* [-BSl]
	Word Formation
-ica 'F.ANIM, DIM' [< BSl & ?Grk]	NEG pronoun < INTRG, e.g., *ni-koj*, *as-kush* [-Grk]
-s- (±-*t*/*d*-) 'verb' [< Grk ± Trk]	numerals 11–19 & 20–90 [-Grk][33]
-ci, etc. 'agentive' [< Trk]	
agentive *-ar* [< Lat]	
	Morphosyntax
GEN = DAT	postposed DEF.ART [-Grk]
short DAT = POSS	aorist/perfect opposition (-BRo)
analytic expression of obliques	complex PREP + finite verb [*za da vidiš*] [-Alb]
analytic comparison of adjectives	status [-Grk][34]
preservation of vocative	
reduplication of the object	
isosyntagmatic prepositional constructions[35]	
absolute relative *deto*, *što*, που, *që*, *că*	
syntactic functions of the definite article	
replacement of infinitive	
future formation based on 'want' and 'have'	
anterior future = irreal conditional	
transformation of preterites in the future	
transformation of "*da*"-imperative in the past	
where = whither [ubi = quo][36]	
repetitions of various types[37]	
perfect in 'have'[38]	

---

33 Some Albanian dialects also show divergence from the Balkan (and Slavic) type.

34 Also called *evidential* (see Friedman 1979).

35 *Isosyntagmatic* is Asenova's term for prepositional constructions and usages which are word-for-word translations among the languages. If it were known that a given construction was translated from one language to another, the term *calqued* could be used, but in the absence of such evidence, the term isosyntagmatic is justified.

36 In his list, Hinrichs 1999a: 432 dismisses this as semantic rather than morphosyntactic.

37 This is given in the text but not in the chart. Examples such as Balkan Slavic *knigi migi* 'books and such,' Albanian *copa copa* 'all in pieces,' etc., can be taken as derivational or morphosyntactic (see §4.3.7 for further discussion).

38 This feature is not given in Asenova's table, but it is adduced by Sh. Demiraj 2004.

2.5   Post-Modern Balkan Linguistics

Strictly speaking, the material covered in this section does not post-date the latter part of the previous section's material. Rather, it represents three types of developments/factors/subjects outside or beyond the traditional field of "classic" Balkan linguistics, the first and second can be conceptualized as internal to Balkan linguistics as a field and as leading to the third, which intersects with the field but is internal to linguistics taken as a whole, i.e., as a diverse but coherent discipline, and leads into the following chapter.

2.5.1 Roms, Jews, and Turks

As already indicated in the Introduction and Chapter 1, Romani (§1.2.3.5), Jewish languages (especially Judezmo), and Balkan Turkic have been excluded from the gaze of Balkan linguistics, either implicitly or – especially during the modern period – explicitly. Weigand's 1895: 78 passing observation that in central Albania many nomadic Roms spoke Romani as their mother tongue while sedentary Roms knew all the Balkan languages *in addition* to Romani (emphasis added) by its very phrasing excludes Romani from the Balkan languages.[39] Sandfeld 1930: 3–4 mentioned Romani and Judezmo in his second footnote for the purpose of excluding them from his study. However, Gilliat-Smith 1915/1916: 68–69 observed that Romani could supply examples for most of Papahagi's 1908 work on Balkan phraseological parallels. Moreover, he (correctly) speculated that Armenian could also be involved.

Turkish has always been kept on the sidelines (as an adstrate) contributing vocabulary, some phraseology and calques, and at most perhaps a verbal category here or a bit of word order there (Friedman 1999d, 2003a: 1–29).[40] It has been studied in a Balkan linguistic context, although almost all such studies are

---

39 The original passage occurs in an account of a journey from Elbasan to Berat with a stopover at the *čiftlik* 'country estate, farm' of Belmak owned by a certain Jussu Bey. As it is one of the few passages in early Balkanistic literature to describe Roms, we cite it here:

> Jedes Gut in dieser Gegend hat seine Zigeuner, die aber nicht nur also Tagelöhner beschäftigt, sondern auch als Bauern angesiedelt sind. Unter den umherwandernden Zigeunern, wie Kesselflickern, Schmieden, Bärentreibern findet man viele, die, aus Rumänien stammend, sich auch der rumänischen Sprache als Muttersprache bedienen, während die Einheimischen außer der Zigeunersprache meist aller Balkansprachen mächtig sind.

> Every property in this region has its Gypsies, who are not only employed as day-laborers but are also settled as peasants. Among the nomadic Gypsies such as tinkers, blacksmiths, and bear trainers, there are many who come from Romania and use Romanian as their mother tongue, while the local (settled) [Gypsies] control almost all of the Balkan languages in addition to Romani.

40 Certain types of verb form usage in Balkan Slavic (see §6.2.5.1) have been attributed to Turkish since Conev 1910/11, and it has also been suggested that the order Genitive-Head and postposed dative possessive clitics have also been influenced by Turkish word order (Friedman 2003a: 1–29; see also §7.4.1.1).

relatively recent. One work in the early modern era of Balkanology that is note-worthy in this regard is Afendras 1968, which contains a section each on the phonology of Judezmo and of Balkan Turkic, with particular attention to different dialects and to the integration of the study of these varieties into the larger Balkan context. Other works can be mentioned on a language-by-language basis. Although Feuillet 1986: 22 puts Turkish on a par with Serbo-Croatian as a peripheral language, Asenova 2002: 23 differentiates the two, pointing out that while Serbo-Croatian elicits mention owing to the fact that some Balkanisms extend into its territory, Turkish "*ne e balkanski ezik, makar če e igral opredelena rolja pri formirajneto na BEC*" ('is not a Balkan language, although it has played a definite role in the formation of the B[alkan] L[inguistic] L[league]'). In neither of these authors, nor in most others after Sandfeld, does Judezmo find any mention, and the dialects of Balkan Turkish are completely absent from all the handbooks and even from Joseph 1983a: 255, which otherwise attempts to be conscientious in its fair treatment of all the Balkan languages.[41]

This is not to say that Romani, Judezmo, and Balkan Turkic have never been studied in a Balkan linguistic context, although almost all such studies are rela-tively recent. For Romani we can cite Kostov 1973, 1998; Uhlik 1973; Joseph 1983a: 252–253; and Friedman 1985a, 2000bc, 2003e as well as Matras 1994a; Sawicka 1997; Boretzky 1995, 1998; and Boretzky & Igla 1999. For Judezmo there are the items cited in Bunis 1981: 36–38 as well as Joseph 1983a: 252–253; Pahmeyer 1980; Kowallik 1992/1993; Gabinskij 1992: 154–173, 1998; Bunis 1999: 60–122 as well as Friedman & Joseph 2014 and Friedman 2013a. Although Skok 1935 treats Turkish in its Balkan context, it is Németh's 1956 classic work that marks the beginning of Balkan Turkish dialectology. However, most of the studies of Balkan Turkic (i.e., Turkish and Gagauz) have been pub-lished in a Turcological context (e.g., Menz 1999, and items cited in Gülensoy 1981 and Tryjarski 1990). Exceptions to this are studies of the influence of the Balkan languages upon local dialects of Turkish such as Jašar-Nasteva 1957, 1969, 1970, 1971/1972, 1986, 1992; Ibrahimi 1982; Friedman 1982c, 2002a, 2003a: 50–83, 2006c; Matras 2003/2004; Tufan 2007; Matras & Tufan 2007; Rentsch et al. 2018, 2020; Winistörfer forthcoming. For Romani and Judezmo, too, there are a number of linguistic works whose frame is Romology or Sephardic or Romance studies, respectively; see, e.g., the bibliographies in Bakker & Matras 2003 for Romani and Bunis 1981: 7–59 for Judezmo. To the extent that these treat dialects spoken in the Balkans, their material can and should be included.

When we seek the reasons for the relative paucity of Romani, Judezmo, and Balkan Turkic participation in the study of Balkan linguistics, a hint at an ideo-logical motivation is found in Sandfeld 1930: 4.[42] In this passage he attributes the commonalities of the classic Balkan languages to the Byzantine Empire and the

---

41 On Balkan Turkish substitution of the optative for the infinitive see Friedman 1982c, 2003a: 63–65, 2006c, and also §7.7.2.1.1.1.6.

42 In our use of "ideological," we have in mind Silverstein's 1979: 193 definition "that ideologies about language, or linguistic ideologies, are any sets of beliefs about language articulated by the

millennium-long domination of the Greek Orthodox Church promulgated by it and subscribed to by most of the speakers of the Balkan languages.[43] The classic Balkan linguistic league is thus cast as a Greek Orthodox, Christian phenomenon.[44] Although some Roms are Orthodox Christians, and moreover the Roms arrived in the Balkans prior to the fall of Constantinople, the majority of Roms in the southern Balkans are Muslims, and in any case, their religion has traditionally been viewed with suspicion by governments.[45] Although Jews were resident in the Balkans before the coming of the Slavs, and in some places even before the Romans, the Judezmo-speaking Jews, like the Muslim Turks, arrived relatively late, and both the timing of their arrivals and their religious confessions would be inconsistent with a theory of the Byzantinogenesis of the Balkan sprachbund. Both the substrate model and the Latinate model of explaining Balkan linguistic phenomena likewise encounter potential difficulty if the languages of Roms, Jews, and Turks are included in the sprachbund, since those explanatory languages are also too far removed in time to account for Balkanisms in these "outsider" languages. If, on the other hand, while acknowledging the antiquity of language contact on the Balkan peninsula, we nonetheless focus on attested phenomena in all the relevant languages, then we see that it is the early modern period, i.e., the Ottoman, not the Byzantine, Empire, that provided the multilingual frame for producing the phenomenon of the Balkan sprachbund as it existed at the beginning of the twentieth century (cf. Asenova 2002: 214&ff. on attested forms of the Balkan future, Joseph 1983a: 179–212, and see §2.5.2).

Aside from the language ideology problem, there is the fact that source materials in and on these marginalized Balkan languages were not being produced at the same rate during the formative period of Balkan linguistics.[46] This was connected in part with the fact that unlike Albanian, Greek, Balkan Slavic, and Balkan

---

users as a rationalization or justification for perceived language structure and use." Cf. also Friedrich 1989: 301, 309; Friedman 1997a.

43 In fact, however, Geg Albanian speakers were mostly Catholic. An exception is made up of twenty-one villages (four of which are now abandoned) in the Reka region of western North Macedonia, whose inhabitants are Orthodox Geg-speakers, sometimes known as *Laramanë*, an adjective whose literal meaning is 'pied, piebald, parti-colored' and sometimes denotes crypto-Christians. We leave to one side that the majority of Albanian speakers converted to Islam, since that occurred after the period to which Sandfeld is referring.

44 Again, leaving Geg to one side, it is worth noting that almost all of Sandfeld's Albanian examples are Tosk.

45 Thus, for example, Evliya Çelebi in the seventeenth century reports that Muslim Roms in Ottoman Turkey were required to pay the taxes expected of Muslims, but also a *haraç*, a tax which was otherwise required only of non-Muslims, because their Islam was suspect (Friedman & Dankoff 1991: 2–3). Nonetheless, Muslim Roms were included in the compulsory exchange of Christians and Muslims between Greece and Turkey as mandated by the Treaty of Lausanne (January 30, 1923). Thus Muslim Romani speakers in Greece (except in Greek Thrace) were expelled to Turkey and Greek Orthodox Romani-speakers in Turkey were sent forcibly to Greece.

46 Although Miklosich 1872–1880 produced important studies of Romani, he did not integrate these with his Balkan linguistic work, nor did those who came after him. Note that in describing these "outsider" languages as marginalized, we are not implying any sort of peripheral status or second-class citizenship of the type posited in Feuillet 1986 and Schaller 1975 for Serbo-Croatian, Greek, and/or Ottoman Turkish. Rather, we are referring to the social status of these languages, which in turn affected their integration into the scholarly literature.

Romance, Romani, Judezmo, and Balkan Turkic dialects were not involved in the production of Balkan nation-states, for which "authentic" materials in national languages were a major ingredient (Bauman & Briggs 2003: 196–225). In the case of Roms, there was no nationalist movement whatsoever at this time. Among Jews, Zionism was dominated by Yiddish-speaking Jews and Judezmo was outside their purview; the debate was between Yiddish and Hebrew. As for Turkish, the purifiers who would create modern Turkish looked to the dialects of Anatolia and the Turkic languages of Central Asia for their sources of purity. It is ironic that just as language ideologies among the non-Turkic peoples of the Balkans looked on contact with Turkish as polluting (as described, e.g., in Kazazis 1977), so, too, Turks looked on their contacts with their non-Turkic neighbors in the same way.[47] Thus, the linguists of the pre-modern and early modern period had at their disposal precisely the folkloric texts that were being collected to build the nations that would work against both the concept and the linguistic convergences of the sprachbund (see §2.5.2; some of these collections actually included materials in other Balkan languages, and in fact those materials have yet to be studied, e.g., Šapkarev 1894, a work currently claimed by both Bulgarian and Macedonian scholars, but in terms of modern dialectology, in Macedonian, which also contains materials in both Aromanian and Albanian).

There are other sociolinguistic differences between the "outsider" three and the "classic" four as well. Thus, for example, unlike the classic Balkan languages, which were the objects of two-way multilingualism, Romani and Judezmo were generally subjected rather to one-way multilingualism. In other words, speakers of the "classic" Balkan languages (and also Balkan Turkish) learned other languages and heard their languages spoken by others. In the case of Romani and Judezmo, however, their positions as languages of stigmatized groups (see Friedman & Dankoff 1991), both of which were strictly endogamous, meant that their speakers were of necessity multilingual but their languages were rarely learned by others (cf. also Rusakov & Eloeva 1990: 8). At the level of the *mahala* ('neighborhood'), of course, children did learn each other's languages, so there were non-Roms who knew Romani and non-Jews who knew Judezmo, and while intermarriages were frowned upon, they did occur (Marushiakova 1992), but all such phenomena were relatively rare vis-à-vis the level of multilingualism among the other Balkan languages.[48] The occurrence of Judezmo words and phrases in folk songs and

47 A similar irony is reflected in the fact that while in the Balkans it is a commonplace to refer to "five hundred years under the Turkish yoke" to account for perceived shortcomings, in Turkey the same shortcomings are explained by five hundred years wasted trying to rule the Balkans (VAF field notes).

48 Weigand 1895: 6 observed that the Vlahs of Bitola understood Judezmo owing to the similarities in their lexicons. He does not indicate, however, that their knowledge was anything but passive. Around the time Weigand was writing, out of a total population of 37,000, 5,500 were Jews, 7,000 Vlahs, 10,500 Turks, 10,000 (Christian) Macedonians, 2,000 Roms, 1,500 (Muslim) Albanians, and 500 Others (Kănčov 1900: 236). There were 3,000 Jews in Bitola in 1943. In March 1943 the Germans and their Bulgarian collaborators rounded up and deported 7,200 Jews from Bulgarian-occupied Macedonia ultimately to the Treblinka death camp in Poland, where all but twelve perished (Todorov 2001: 8–9). They also deported 4,143 Jews from Bulgarian-occupied Thrace, as well as Jews from the Torlak regions of Serbia (also occupied by Bulgaria), and used Bulgarian

folk poetry or of Romani words in slang or secret languages (e.g., Cvetkovski 1988: 190; Jašar-Nasteva 1987) does not contradict this principle but rather is an indication of the relative rarity of multidirectional multilingualism (as opposed to unidirectional multilingualism) in these languages. Relevant too are codeswitching phenomena in Macedonian folk tales, in which Jews speak Turkish rather than Judezmo (Friedman 1995b). Of these three languages (Romani, Judezmo, and Balkan Turkish), Romani is the most viably Balkan in terms of surviving multilingualism, despite the depredations of the Holocaust, which targeted Roms as well as Jews (Kenrick & Puxon 1972). The Jewish communities of the Balkans were almost entirely destroyed by the Nazis and their collaborators during World War Two, and those that survived generally emigrated to Israel. Balkan Turkish is steadily losing ground to both migration and to other assimilatory factors (e.g., the dominance of Albanian among Muslims in North Macedonia and the spread of Standard Turkish as a result of increased access to Standard Turkish media). These changes in the linguistic environment from the days of Sandfeld bring us to the interrogations discussed in §2.5.2.

2.5.2 Questioning the Question

In §2.5.1 we discussed three speech communities whose languages or dialects have been traditionally excluded from Balkan linguistics. In this section, we examine various developments problematizing Balkan linguistics as a discipline. Regardless of whether one views the Byzantine or the Ottoman Empire, or both, as being formative of the Balkan sprachbund, the fact remains that the political structures that provided the frame for Balkan multilingualism had ceased to exist by the second decade of the twentieth century.[49] Following a course of events begun in the late eighteenth and early nineteenth centuries, the transformation of empire into nation-state and the ideology of the nation-state as a homogeneous unit with fixed borders (and a single "national" language) created conditions that were antithetical to Balkan multilingualism. Ideologies of purity and the notion of a standard language based on the vernacular but at the same time bounded and closed, excluding so-called "folk speech," finds its analogy in other nineteenth century ideologies that appropriated "folk" material in order to refashion it as a vehicle for modern urban (bourgeois) acculturation (Baumann & Briggs 2003: 163–196). The ideology of purification that would make way for a new hegemonic hybridity had its most extreme realization in large-scale movements of

Jewish slave labor to construct some of the necessary rail lines (see https://en.wikipedia.org/wiki/The_Holocaust_in_Bulgaria, which contains both good information and references to other resources; see also Comforty 2021).

49 Although Masica 2001 highlights the need to differentiate political from linguistic areas, in the case of the Balkans (and elsewhere) the patterns of communication established by politically determined freedom or direction of movement can indeed result in isoglosses as is the case in the Rhenish fan (Bloomfield 1933: 341–345) and as may be the case with the major isoglosses separating northeastern Yiddish from southern Yiddish (Jacobs 2005: 62).

populations.[50] Such movements began with Muslims leaving former Ottoman territory as the empire contracted (McCarthy 2002: 135–151). In the aftermath of the Balkan Wars, the World War One treaties of Neuilly (November 27, 1919) and Sèvres (August 10, 1920) made provisions for people to "opt" for the nationality of a state other than the one in which they found themselves owing to the establishment of frontiers, and required them to move to that state if they did so opt (see Ladas 1932 on Greco-Bulgarian "voluntary" exchanges).[51] In the wake of the Greco-Turkish war, the Treaty of Lausanne (January 30, 1923) went a step further and mandated "a compulsory exchange of Turkish nationals of the Greek Orthodox religion established in Turkish territory, and of Greek nationals of the Muslim religion established in Turkish territory."[52] The majority of rural immigrants from Turkey were settled in Greek Macedonia and contributed greatly to the Hellenization of the region by adopting Greek, even if their first language had been Turkish or some other language, and by pressuring local non-Greek inhabitants to leave (Pentzopoulos 1962: 27–48, 125–140). The aftermath of the Greek Civil War (1946–1948) resulted in the exodus or expulsion of even more non-Greek speakers (mostly Albanian and Macedonian) from northern Greece. Communist policies in the Balkans after World War Two, especially in the 1950s, also encouraged the emigration of Muslims to Turkey and Greek-speaking Sarakatsans to Greece (Ellis 2003: 43–63; Nedelkov 2011). Thus, not only was free communication over a large part of the Balkans impeded or restricted, but the attempt to build homogeneous nation-states also discouraged the development of multilingual situations.

50 It is important to note that significant population movements are an essential part of the history of the Balkans (and many if not most other regions). From the invasion of the Indo-European peoples whose languages were ancestors of Modern Greek and Albanian as well as the vanished languages and whose languages replaced those that had preceded them (apparently both Indo-European and non-Indo-European), to the coming of the Romans along with internal shifts and strife (see Papazoglu 1978 on the Triballi (pp. 58–86), Autariatae (124–129), Dardanians (186–225, 256–269), Scordisci (345–389), and Moesians (402–409, 430–437) in the central Balkans), to the so-called Great Movement of Peoples of the early Middle Ages that, among other things, brought the Slavs to the Balkan Peninsula, to population shifts as various groups sought to expand their territory or to get away from the violence of the expansionists, to Byzantine and Ottoman Imperial policies of moving whole groups into or out of the Balkans (cf. Soviet policies in the Caucasus and elsewhere from the consolidation of power through World War Two), to groups at various times and for various reasons migrating to seek better living conditions (the Judezmo diaspora is well known, on the Serbs, for example, see Ivić 1971, and on the eastern Albanian diaspora see Friedman 2004a: 59–155, also Liosis 2021), there was rarely a period devoid of some sort of population movement (cf., e.g., the articles in Schmitt 2016 for the Ottoman period). Nonetheless, those shifts that took place from the decline and fall of the Ottoman Empire to the aftermath of World War Two were, unlike previous shifts, directly connected with the modern ideology of the homogeneous nation-state or a combination of politically determined economic conditions within a nation-state.

51 In the case of Turkish subjects who found themselves outside the borders of Turkey, the provision was that they could opt for Turkish nationality and move to Turkey. In all other cases, it was stipulated that persons opting to change nationality (i.e., citizenship) – and therefore move – were a minority in the state in which they found themselves and a majority in the state to which they moved (Treaty of Sèvres Articles 123–126, Treaty of Neuilly Articles 40, 45).

52 Treaty of Lausanne, Article 1. The Greek Orthodox Christians of Constantinople and the Muslims of western Thrace were excepted, as were the Çams (Albanian-speaking Muslims) of Epirus (Ladas 1932: 380–387).

It is in this connection that Topolińska 2000 makes the claim that the Balkan sprachbund as such no longer exists except as an historical artifact – rather like a given stage in the development of a genetic/genealogical family. While it is indeed true that the creation of national boundaries has broken up the larger unity that constituted the Balkan sprachbund as such, that the standardization of national languages and the concomitant effect of mass literacy and education have had a similar effect on linguistic development, and that, moreover, the rise of English as the international lingua franca has taken the place of local languages in contact situations, nevertheless, the same type of multilingualism with the same Balkan languages continues to exist at the communal level in all the Balkan countries, even those that claim that it does not exist.[53] Despite increasing ethnic isolation (e.g., Icevska & Salihu 1998) and aggressive assimilatory policies in various Balkan countries (various Human Rights Watch/Helsinki Watch reports attest to this), all of which work against neighbors knowing the language of neighbors, there is still much that can be done to investigate both the remnants and the continuations of the Balkan sprachbund.[54]

In such a context, it is also important to reexamine dialectal accounts that homogenize and totalize larger areas, thereby erasing (Gal & Irvine 1995) import-ant sources of variation within them. A single example to illustrate our point is the usual account of Geg Albanian that describes it in terms of expressing the future by means of the conjugated present of 'have' + infinitive, which is characteristic of Romance, rather than future expressed by the invariant marker derived from 'want' + subjunctive/conjunctive characterized as Balkan. In fact, however, there is considerable diversity within Geg, and even in Tosk, when Arbëresh is taken into account. Thus, for example, in Kelmend in Northwest Geg, the 'want' future is used in speculations (Shkurtaj 1975: 55). Further west, along the left bank of the river Buna, only the 'want' + subjunctive future occurs (Gjinari 1971: 252). A similar situation obtains to the southwest, in Puka (Xh. Topalli 1974: 316), which is transitional between the northeast and the northwest, although its center of gravity is Shkodër in the northwest. However, Shkrel, southeast of Kelmend, uses only 'want' + subjunctive (but also *tash* 'now' + progressive *po* + present indica-tive; Beci 1971: 298). In the southern part of Northeast Geg, e.g., Has (Gosturani 1975: 237) as well as the Presheva (Srb Preševo) valley (Badallaj 2001: 178), the future in 'have' is limited to a sense of obligation while 'want' + subjunctive is more volitional. In Upper Reka, the future in 'have' + infinitive has been

---

53 The worst offender in this matter is Greece. Šmiger [Schmieger] 1998:20–21, for example, thanks the Greek police for teaching him the value of back-up copies. When Schmieger was doing his fieldwork in the Macedonian-speaking village of Nestram (Grk Nestório) in Greek Macedonia, the local police confiscated and destroyed his tapes, unaware that their victim had made back-up copies.

54 Friedman 2003b observes that in the 1994 Macedonian census more people declared knowledge of English than knowledge of a second Balkan language (in the case of Macedonians) or a Balkan language other than Macedonian (in the case of other groups). In that same article, he notes a significant change in the direction of consistency of declared nationality and declared mother tongue in the four decades between 1954–1994, itself an indication of the hardening of ethnolin-guistic boundaries. Still, as the articles in volumes such as Sobolev 2021c demonstrate, Balkan multilingualism as such continues into the twenty-first century.

completely replaced by 'want' + subjunctive (Haruni 1994: 76). South of Has and west of Upper Reka, in Luma, the two types of future are in competition, but the 'want' type predominates (Hoxha 1975: 165, 1990: 136). West of Luma, in Mirdita, the 'want' future is regular and the 'have' future is rare (Beci 1982: 84–85). Similarly, in Tuhin (Mac Tuin), southeast of Upper Reka and northeast of Kičevo (Alb Kërçova) in RN Macedonia, the 'want' future (with indicative) predominates, although 'have' + subjunctive also occurs, as it does in the Tosk dialects of Italy (Arbëresh) with relics also in Labëri (Totoni 1971: 73). In this region of RN Macedonia, as in transitional dialects such as Shpat, as well as Luzni (southwest of Peshkopi), the 'have' + infinitive future uses *për* + verbal noun *(= të* + participle) rather than *me* + participle. The *me* + participle construction is extremely rare in Tuhin, although its opposite (*pa* + participle 'without VERB-ing'; cf. *me* 'with' vs. *pa* 'without') is quite common (Murati 1989: 41, 44; Çeliku 1971: 230; Beci 1974: 250). In general it is worth noting that in Albanian *pa* + participle often has a literal meaning, e.g., *pa u lagur* (lit., 'without becoming wet') 'unharmed,' but it can also have a temporal nuance of 'before,' e.g., *pa u tharë (mirë) gjaku* 'before the blood was (even) dried' (lit., 'without being-dried (even) the-blood'; cf. Newmark 1998: s.v. *pa*). Thus, while Geg does have conjugated 'have' + infinitive in contexts where Tosk uses invariant 'want' + subjunctive, the characterization of Geg being opposed to Tosk in a simple binary manner in this respect fails to capture the complexities of Geg usage. In fact, Southern Geg goes with Tosk (including Arvanitika), while Northern Geg and Italian Tosk (i.e., Arbëresh) are linked by the use of 'have' as the future marker. Even when standard languages are taken as the source, a simple characterization such as "loss of the infinitive" does not capture the real complexity, as was emphasized by Joseph 1983a.

The typical characterization of the type 'infinitive loss' brings us to another point made by Topolińska 2004, who notes the need for Balkan linguistics to reorient not just its terminology, but the manner of investigation:

> It seems that the classical inventory of Balkanisms, such as that which we know from synthetic studies, contrastive grammars, etc., was created on the "negative" principle: it presents a list of "deficiencies" of the Balkan linguistic systems as compared to the prototypical Indo-European, or – as in the case of Macedonian and Bulgarian – in relation to Common Slavic. We should reformulate and enrich that inventory starting from the positive, constructive, functional point of view. Thus, for example, instead of regretting the 'loss of the infinitive', we should emphasize the strict grammaticalization and the enlarged zone of use of the subjunctive, instead of speaking of the 'loss of declension', we should speak of coding the case-relationships on the morphosyntactic level, etc. We should speak of the Balkan way of organizing the /-factive/ part of the verbal system, of the Balkan network of synsemantic verbs and adverbial particles, etc. It would be not only a terminological shift, but also a new and constructive approach for further research.

Here Topolińska is pointing to the origins of Balkan linguistics in Western European and Slavic linguistic contexts, in which the languages of the scholars in question (from Kopitar to Trubetzkoy, Sandfeld, and beyond) were taken as

a norm and the Balkan languages taken as deviations from such a norm. Not only is it the case that Balkan linguistic developments can and should be viewed in their own context as communicatively motivated expansions of certain grammaticalized functions, but moreover, in certain instances, such as the expansion of analytically expressed case relations, a certain Slavocentric bias needs to be overcome. After all, it is indeed the case that in Macedonian, Bulgarian, and Torlak BCMS the expansion of analytically expressed case relations and the elimination of inherited synthetic expressions (cases) have gone further than elsewhere in Slavic;[55] it is also true that Balkan Romance has been uniquely conservative among the Romance languages in preserving traces of synthetic case, and both Greek and Albanian, as well as Romani and Balkan Turkish, have likewise failed to expand analytic expressions of case to the same extent as Balkan Slavic.[56]

This in turn brings us to an important but generally neglected field of Balkan linguistic investigation, namely the interrelationship between convergence and divergence and the need to distinguish them carefully. Friedman 1983, using Macedonian and Albanian data, noted the fact that superficial resemblances some-times mask underlying differences of structure. Friedman 1978 likewise observed that superficial resemblances between Bulgarian and Turkish could be explained by contact-induced convergence rather than simple borrowing. Fielder 1999 has expanded on this. (See §6.2.3 for details of these examples.) The different types of multilingualism mentioned above (§2.5.1) also involve differences between contact phenomena and boundary maintenance. This can be seen in the Macedonian dialects of Balkan Romani, where phonological conservatism seems to serve as a marker of ethnolinguistic boundary in the face of syntactic conver-gence (Friedman 2000b), and within syntax the noun phrase seems more resistant to contact phenomena than the verb phrase (Friedman 2000c). Moreover, within the verb, tense/aspect is more conservative than mood (Friedman 2001a; Matras 2002: 151–165).

The profound changes that took place during the course of the twentieth century completely altered the networks of contacts that produced the Balkan sprachbund even as that sprachbund was being theorized and studied in modern linguistic frameworks. The fixing of borders that reduced communication and disrupted traditional patterns of multilingualism, together with the creation and

---

55 And we should note that, strictly speaking, even within Balkan Slavic there is variation between the preservation of traces of inherited nominal inflection at the margins and complete elimination in the center. See §6.1.1 for details. Noteworthy too is the simplification of cases in adjectives (as opposed to other nominals) in southern Montenegrin (Stevanovič 1935: 76–77).

56 In the case of Judezmo, it would appear that the relevant Western Romance changes had already been completed by 1492, while in the case of Romani in North Macedonia the tendency toward calqued analytic encoding of nominal relations is particularly strong. Similarly, Aromanian dialects in close contact with Macedonian such as Ohrid tend to eliminate what is left of morphological case distinctions (Markovikj 2017: 51). Likewise, in Meglenoromanian, which has been heavily influ-enced by Macedonian, nominal case relations are expressed entirely analytically (except for the village of Ljumnitsă (Grk Skra); Atanasov 1984: 195–197). We can also note here that Albanian synthetic declension is somewhat agglutinative insofar as some case markers do not differ for gender.

spread of standard languages that frequently relegated features typical of the sprachbund to lower or archaic registers, has created a situation in the twenty-first century in which the Balkan sprachbund is, in certain respects, an historical artifact. At the same time, the rise of English as a global lingua franca of unprecedented proportions has altered patterns of borrowing and calquing, and, potentially, of code-copying, and possibly even metatypy and fusion (Johanson 1992, 2002, 2023; Ross 2001; Matras 2000).[57] In its Balkan context, English has become the Turkish of the twenty-first century: it is the major source of loanwords and new calques, and speakers of majority languages in the Balkan nation-states are more likely to know English than any of the national or minority languages of their immediate neighbors. Nonetheless, the previous exclusion of marginalized groups (see §2.5.1), the continuation of the sprachbund at the local level, and new approaches that take into account the greater array of dialectological materials available and also new theoretical approaches (see §2.5.3) all promise new and original understandings of the area in particular and contact phenomena in general. Moreover, as we see below, even the facts of the Balkan sprachbund as it has been theorized and described thus far deserve to be better known.

2.5.3 The Balkan Other and Other Linguistics: Areal, Contact, Typological, Ecological, Eurological

In this chapter we have been concerned with the history of Balkan linguistics as an independent field of investigation and with directions for further development in and of that field. As a branch of the larger discipline of linguistics, Balkan linguistics constitutes an oft-cited but little-utilized example of contact-induced language change (cf. Anttila 1972: 172–175; Campbell 1998: 287–310; Croft 2003: 24, 34, 36; Crowley 1992: 259; Hock & Joseph 2019: 350–356; Hock 1988: 259; McMahon 1994: 248–250; Myers-Scotton 2002: 174; Thomason & Kaufman 1988: 88, 95–96, 147; Thomason 1997: 105–109; Trask 1996a: 315–317; Weinreich 1968: 31–32, 50, 113; and Winford 2003: 8, 13, 64, 70–74; Heine & Kuteva 2005: 187–199 spend more time than most, but in a different sort of theoretical framework). In the next chapter, therefore, we survey relevant theoretical developments in the discipline of linguistics as a whole. Of particular concern to us is the contrast between Areal (now Contact) linguistics versus typological explanations (e.g., Hamp 1977a), the ecological approach to choices among competing alternatives to communicative efficacy (Haugen 1972; Mufwene 2001a), the epidemiological approach to the spread of language change

---

57 We can also note here the role of new international entities such as the European Union (EU). Nonetheless, if anything, the EU is an aid to the rise of English, despite French and German desires. We could also note altered patterns of migrant labor, which during the Ottoman period was conducted mainly within the borders of European Turkey but has shifted to wealthy Western European and anglophone New World countries. However, this pattern had already begun with the upheavals of the decline of the Ottoman Empire (K. Brown 1998, 2015).

(Enfield 2003), and the delimitation of areas such as the Balkans, Europe (the Eurotyp project, Reiter 1994; Haspelmath 1998; Heine & Kuteva 2006), Eurasia (Masica 1976, 2001), and beyond (e.g., Nichols 1992). It is also here that we discuss the principal concepts and methodologies that inform our approach to the field throughout this book.