Introduction

Jazicite se bogatstvo [Macedonian]
Limbali san avuti [Aromanian]
Limbili să vut [Meglenoromanian]
Ezikăt e bogatstvo [Bulgarian][1]
'Languages are wealth'
Sa më shumë gjuhë të dish, aq më tepër vlen [Albanian]
So pobuter čhiba džaneja pobuter vredineja [Romani (Arli)]
Što više jezika znaš, to više vr(ij)ediš [Serbian/(Bosnian/Croatian//Montenegrin)]
'The more languages you know, the more you are worth'
Bir lisan, bir insan; iki lisan, iki insan [Turkish]
'One language, one person; [one who knows] two languages [is worth] two people'
[also cited in Turkish by Judezmo-speakers]
Cu fiecare alta limba traieşti o nouă viaţă [Romanian]
'With each new language you live a new life'
Όσες γλώσσες έχει ο άνθρωπος, τόσα κεφάλια έχει [Modern Greek]
'As many languages as a man has, he has that many heads'[2]
Quintus Ennius tria corda habere sese dicebat quod loqui Grecae et Osce et Latine sciret. [Latin; 2nd cent. CE]
'Q. E. used to say that he had three hearts, because he knew how to speak Greek, Oscan, and Latin' – Aulus Gellius, *Noctes Atticae*, 17.17.1

The Balkan peninsula is home to a number of languages whose interrelationships present various elements of interest for questions of language contact, language change, areal linguistics, and sociolinguistics. Between 1975 and 2018, no fewer than thirteen book-length general treatments of the Balkan languages were written: Schaller 1975; Solta 1980; Banfi 1985; Feuillet 1986; Desnickaja & Tolstoj 1990, 1998; Reiter 1994; Steinke & Vraciu 1999; Hinrichs 1999a; Asenova 2002; Sh. Demiraj 2004; Feuillet 2012; and Fiedler 2018.[3] All of these works build in

1 The Bulgarian means 'the language is wealth' but 'language' here means 'any language' and the connotation is the same as in the other expressions.
2 This proverb is not well known to contemporary Greeks, but it occurs in Politis 1899, a collection of proverbs, and thus presumably was current at least into the nineteenth century. We note that some Greek speakers in Himara in southern Albania have given a similar proverb-like opinion (BDJ field notes, September 2014).
3 Cf. also Civ'jan 1979, 1990, and Nikolaeva 1996. Schaller's 1977 Balkan bibliography, which stops at 1975 and contains 1,536 items, while not a handbook and already considerably out of date given subsequent progress in the field, is nonetheless the only reference work of its kind and thus is worth mentioning here. We can note that Asenova's 1983 Balkan linguistic bibliography has 1,996 entries for the period 1966–1975 alone. Alexander 2000a has over 1,200 items for South Slavic dialectology

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

some way on the classic 1926 study by Kristian Sandfeld, better known in its 1930 French version, whose synthesis of the previous century's research combined with his own analyses is considered the founding work of the modern field of Balkan linguistics – the subject of this book. By *Balkan linguistics* we mean the study of the special relationship that obtains among the languages of Southeast Europe that form the Balkan linguistic league.[4] It was the grammatical convergences that characterize the Balkan languages that led Trubetzkoy 1923, 1930 to theorize, for the first time in modern, explicit terms, the ways that linguistic structure can change not only owing to divergence but also as a result of convergence in multilingual contact, and that therefore in addition to the concept of language family – a group of languages whose resemblances result from descent from an earlier common form – there is also the linguistic league (in German, *Sprachbund*), i.e., a group of languages whose resemblances result from mutual influence.[5]

Despite the debate that emerged after 1989 around the use and definition of the term *Balkan* (cf., e.g., Bakić-Hayden & Hayden 1992; Bakić-Hayden 1995; Todorova 1997; Aronson 2007; and cf. also Friedman 1997a: 36, 2000a: 142, 2011a; see Račeva 1996 for an interesting etymological note), there are four living

and Friedman 2013a is a selective, annotated Balkan bibliography. Rosetti 1985 is a useful collection of the author's articles, but not a handbook. Haarmann 1978, even though entitled *Balkanlinguistik*, is not a general survey: the first volume is a lexicostatistic account of Balkan Latinate vocabulary, the second is on Moldavian sociolinguistics. Cyxun 1981 concentrates on Balkan Slavic. We must also mention here Tomić 2006, which is marred by a plethora of errors, inaccuracies, misleading formulations, and inadequate citations (see Sims 2008; Joseph 2011d).

4 We thus exclude from general consideration here the many shared aspects of Balkan material culture – interesting though they may be in their own right – including such spheres as apiculture, viniculture, viticulture, clothing, tools, etc. (see, e.g., Civ'jan 1990; Mladenova 1994; 1998; Ostuni 1986; Sujecka 2012; Svešnikova et al. 2007 for scholarly treatments of some of these areas), and spiritual cultural domains as defined by various religions and other belief systems. These common cultural traits do help define a "Balkan" area in other than linguistic terms and thus may be a useful and important complement to what we discuss here, but they raise issues that extend beyond our present scope. Some discussion of the linguistic dimensions of shared culture, in the form of loanwords, is given in Chapter 4. We can also note here Sobolev 2005–2013, Domoselickaja 2012, and Domoselickaja & Sobolev 2018, which cover a variety of cultural and lexical fields.

5 Simpson 1994: 210 takes as the original formulation of areal linguistics Schleicher's 1850: 143 statement, which, using the ideological models of his day, refers to the Balkan languages as having "corrupted" one another despite belonging to different families (i.e., groups). As Sandfeld 1930: 12 pointed out, it is Miklosich 1862 that contains the first explicit list of Balkanisms, i.e., some of the grammatical traits common to the four classic Balkan language groups (types of future formation, uses of subjunctive clauses, etc.). The term *Balkanism* itself first occurs in Seliščev 1925. Nonetheless, it was Trubetzkoy in 1923 (in Russian, *pace* Thomason 2001: 99) and again in 1928 (Trubetzkoy 1930), who first explicitly formulated the linguistic league as a theoretical concept. The German term *Sprachbund* arguably has been borrowed into English-language academic usage just as have terms like *gestalt* (plural, both *gestalts* and *gestalten*). Therefore, throughout this work, *sprachbund* is not capitalized or italicized except when it is being cited as a German word (cf. Dreyer 2019: 74–75) or used in an unaltered quotation; we also use the plural *sprachbunds* (like *gestalts*) rather than *sprachbünde* (like *gestalten*).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

Indo-European language groups which, having come into contact for centuries or even millennia in Southeastern Europe, are universally recognized as containing members that constitute the classic Balkan sprachbund: Balkan Albanic, Balkan Hellenic, Balkan Romance, and Balkan Slavic (see Chapter 1, passim). We use the adjective *Balkan* in each case, since there are members of each of these groups spoken outside the Balkans.[6] To these four can be added Balkan Turkic, which is generally treated as a contributor to but not a *participant* in the Balkan sprachbund (but cf. Friedman 1982c, 2003a: 50–83), and Balkan Indic, i.e., the Romani dialects of the Balkans, which, until Hinrichs 1999a, were never accorded more than passing mention in the general handbooks and were rarely treated in specialized studies (but see to the contrary, e.g., Gilliat-Smith 1915/1916; Kostov 1973, 1998; Uhlik 1973; Joseph 1983a: 252–253; Friedman 1985a, 2000b, 2000c; Matras 1994a; Boretzky 1995, 1998). We can also mention here Balkan Judezmo (sometimes called Ladino, Judeo-Spanish, etc.; here we accept the judgment of Bunis 2018, cf. also Harris 1982; Yenisoy 2015), which is Romance but generally not treated with Balkan Romance languages (but cf. Joseph 1983a: 252–253; Kowallik 1992/1993; Gabinskij 1992: 154–173, 1996, 1998; Desnickaja & Tolstoj 1998; Friedman 2006a, 2021; Friedman & Joseph 2014, see §1.2.3.3).[7] We discuss the specifics of our definition of *language* and of *Balkan language* in §1.2.

In the first chapter we also give a general survey of the relevant aspects of Balkan geopolitical, cultural, and linguistic history and discuss other extralinguistic factors relevant for the linguistic situation. In Chapter 2, we present a historical survey of the field of Balkan linguistics, and in Chapter 3 we discuss the key methodological and theoretical issues relevant for Balkan linguistics as a specific manifestation of complex language contact. Of particular significance here are studies of other proposed linguistic areas such as Amazonia (Aikhenvald 1999, 2002), Araxes-Iran (Stilo 2018, 2023), the Caucasus (Catford 1977; Klimov 1978; cf. also Friedman 1996a; Tuite 1999; Comrie 2008), Ethiopia (Ferguson 1970; Zaborski 2010 and references therein, but see also Tosco 2000), Mainland Southeast Asia (Enfield 2003, 2005, 2017), Meso-America (Campbell et al. 1986; Suárez 1983), the Northwest Coast of North America (Beck 2000), South Asia (Emeneau 1980; Gumperz & Wilson 1971; Masica 1976, 2001), parts of Australia (Dixon 2002: 668–689; Heath 1978: 140–143) and of Papua New Guinea (Ross 1996, 2001).[8]

6 In some cases, the extra-Balkan members of a given language group emigrated from a Balkan context (e.g., the Arbëresh in Italy) while in others the relatives never entered the Balkans (e.g., Oscan for Italic or Belarusian for Slavic). The relevance of extra-Balkan members of groups that have Balkan representatives is discussed in §1.2 and throughout this work as relevant, but cf. especially §7.7.2.1.5 on the geography of developments with nonfinite complementation.

7 As noted in the User's Guide (p. xxiv), here and throughout this work, mention of "§X.Y . . . " refers to Chapter X, section Y, with any other digits specifying the relevant sub-sections and sub-sub-sections (etc.) within that chapter and section.

8 Cf. also the proposed Circum-Baltic area (Dahl & Koptjevskaja-Tamm 2001) and also case studies in Matras et al. 2006; Aikhenvald & Dixon 2007; Hickey 2017; and Adamou & Matras 2021. For a critique of some approaches treating all of Europe as a linguistic area, see Friedman 2008a. Hamp 1989a, however, gives a genuinely nuanced approach to contact-induced change across Europe,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

The Balkans represent not only the most studied such case but also the most studiable, in that of all the sprachbunds that have been discussed in the literature, the Balkans offer the greatest amount of and the longest time-depth for information on the linguistic history of the area, the social history of the peoples in the area, and relevant reconstructible linguistic prehistory. Regarding this last point, we note the considerable evidence for earlier linguistic stages that comes from the attestations in texts and inscriptions for the ancestral languages of four out of the five Indo-European groups (Ancient Greek for Hellenic, Latin for Romance, Old Church Slavonic for Slavic, and Sanskrit for Indic), leaving only the ancestor of Albanian as problematic and without any textual attestations above the level of the word (but cf. Hamp 1982). Further, the Comparative Method offers another point of entry for Balkan linguistic prehistory, working with evidence from related languages elsewhere in the families.

These three chapters together constitute the first part of the book. The second part, comprising Chapters 4 through 7, treats Balkan sprachbund phenomena in the lexicon, phonology, morphology and morphosyntax, and syntax, respectively. Pragmatics, and semantics are treated, where appropriate, throughout these chapters, and especially in Chapter 4 for lexical semantics. In all of these chapters, we cover both features widely discussed in the literature and some brought to light for the first time here or only given brief mention elsewhere. Although, as we noted previously, it was convergences in grammatical structure that led to the theorization of the Balkan sprachbund as a linguistic entity whose origins are due to contact-induced change, convergences at other levels are also relevant to a comprehensive study of the field. It is also worth noting here that divergences – and particularly superficial convergences that at a deeper level represent divergent structures (cf. Friedman 1978, 1983; Sobolev et al. 2018) – are equally relevant in a comprehensive treatment of a linguistic area such as the Balkans, and they are noted where relevant.

A major issue in understanding the Balkan sprachbund is causation. The causes of convergence and divergence in the Balkans are discussed throughout the presentation of particular features, regarding language contact and other factors, including sociolinguistic phenomena such as multilingualism (both productive and passive, i.e., receptive; see §3.2.1.1), codeswitching, the roles of politically and culturally dominant or prestigious and also subordinate languages, i.e., superstrate, substrate, and adstrate, also acrolectal, mesolectal, and basilectal factors and structural mechanisms (the role of prosody, selection among competing features, internal motivation versus borrowing). The final chapter, Chapter 8, provides a summation of the main issues pertaining to causation, and offers a careful

which he characterizes as "a spectrum of differential bindings, a spectrum that extends in different densities across the whole of Europe and beyond."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

assessment of the sort of construct the Balkan languages represent, by way of problematizing the notion of "sprachbund."[9]

Given the large number of surveys of Balkan linguistics cited earlier, the reader might question the justification for another work on this topic other than the fact no such work has ever been available in English. There are, however, a number of reasons why the present book proposes to contribute something new to the ongoing study of the Balkan languages.

To begin with, we can note that certain specific contact phenomena in the Balkan languages, commonly referred to as *Balkanisms*, have not been exhaustively investigated or have been described in scattered places. Thus, for example, observations by Joseph 2001b on negation and Friedman 2003c, 2014a on indefinite marking and the narrative imperative, as well as other data, have yet to find their way into general accounts of Balkan linguistics but are given voice here. We can also include here more local phenomena, e.g., the Albanian-based admirative construction in the Fărshălots (Frasheriote) Aromanian dialect of Bela di Suprã (Mac Gorna Belica, Friedman 1994b) or the Aromanian-influenced marking of direct objects with prepositions in southwestern Macedonian (Koneski et al. 1968; Koneski 1981), which are also generally absent from the handbooks (see §3.5 and §6.1.1.1.2).

Another reason for this volume is the emergence of contact linguistics, to which areal linguistics can be construed as a contributing field, as an increasingly popular focus of scholarly inquiry (of the several handbooks on these fields, Mufwene & Escobar 2022 appeared just as this book was going to press). From the time of Sandfeld and Trubetzkoy to the present, the majority of Balkan linguistic handbooks have remained within the traditional paradigm as it was first defined. This paradigm, which basically concentrates on a combination of specific morphosyntactic, lexical, and, rarely, phonological, features, lends itself to a kind of listing approach (cf. Lindstedt 2000; van der Auwera 1998; Campbell et al. 1986), which has the disadvantages of completely eliding the social mechanisms of linguistic change and failing to address how the relevant elements fit into their respective languages, whose systems often show underlying divergences masked by surface convergences (cf. Friedman 1978, 1983). This last point is itself an important problem in the study of the nature of language contact. Of all the handbooks that have appeared since Sandfeld's important works of 1926 and 1930, Asenova 2002 represents the most substantive advance within the traditional paradigm. Those works that have attempted

9 The effects of post-Ottoman modernity on the Balkan sprachbund as well as questions of perception and continuity, the rise of modern nation-states and their linguistic vehicles of governance (power), i.e., the standard languages, including such related issues as national consciousness, ethnic identity, language planning, and linguistic purism, are treated in numerous individual studies (e.g., Friedman 1996b, 2003b). A comprehensive treatment, however, remains a separate project for the future and is beyond the scope of this book. We can note here, however, based on field research we have conducted in the twenty-first century, that Balkan multilingualism of the type that characterized the Ottoman and early post-Ottoman periods is still to be found at the local level in various locales and to varying degrees in every modern Balkan nation-state (see Friedman 2011b).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

different approaches, while useful and interesting, have either been unintegrated collections of articles on specific languages or topics (Desnickaja & Tolstoj 1990, 1998; Hinrichs 1999a; Sobolev 2021c) or have basically abandoned the areal model (Reiter 1994).

Thomason & Kaufman 1988 took the areal model a step further in uniting various types of contact-induced language change (including creolization and pidginization) into a single model of contact linguistics in opposition to the genetic or genealogical model. Their explicit unwillingness to treat multilingual contact situations in general (Thomason & Kaufman 1988: 95), however, and the Balkans – which they call "the world's most famous contact situation" – in particular, while understandable within the parameters of what they were trying to achieve, nevertheless leaves the phenomenon uninterrogated. Having observed that "Sprachbund situations are notoriously messy" (Thomason & Kaufman 1988: 95–96), they devote only a paragraph and two or three comments in passing to Balkan phenomena. In subsequent work on contact linguistics, the Balkans have been similarly cited as well known and then passed over. Typical in this respect is Myers-Scotton's 2002: 174 paragraph on the Balkans that simply lists ten features culled from Hock 1988: 286–290, some diagnostically Balkan, some not.[10]

The present study of the Balkan linguistic league, building as it does both on the achievements of traditional Balkan linguistics and on modern approaches to language contact in its social setting, aims at new answers to questions of how language contact works in complex multilingual situations in general and in the Balkans in particular. In addition to the importance of contact phenomena in shaping the structures of the Balkan languages, however, we also address the question of limitations, i.e., those portions of the grammar that are resistant to contact-induced change or where apparent surface resemblances mask differences at other levels (cf. Friedman 1978, 1983, 2001c). In so doing, unlike various scholars (e.g., Andriotis & Kourmoulis 1968; A. Belić 1936; Birnbaum 1965, 1966; Minčeva 1987: 139–149; S. Mladenov 1939; Reiter 1991, 1994; Steinke 1983), we are not attempting to minimize the importance of contact, but rather to contribute to the discussion of the nature of constraints (cf., e.g., Myers-Scotton 1993a,b, 2002, but to the contrary Thomason & Kaufman 1988; Friedman 2007a).

Beyond modern approaches to contact linguistics, which, as we have noted, have eschewed dealing with the complexities of Balkan multilingual contact, the Balkan languages and the interesting linguistic phenomena they present are mentioned in other works on general linguistics, but often with erroneous and even comical results, as with Hopper & Traugott's 1993: 210 (also 2003: 213) reference to Sandberg [*sic*] 1930, or Croft's 2003 reference to articles and definiteness in the Balkans that is stated in such a way that Greek, in the

10 Her inclusion of Croatian, no doubt based on Hock's Serbo-Croatian, is also problematic; see §1.1fn.5 and §1.2.3.4 for discussion.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

Balkans itself, is a counter-example to his generalization. Similarly, many general linguistic surveys cite Balkan examples, but by their very nature such surveys cannot engage in the type of detailed analysis that gives a nuanced picture; the result is a presentation of standard, often prescriptive examples that fail to capture the full range of complexity that the facts demand.[11] And, there are general works as well that devote (necessarily) limited discussion to the Balkan sprachbund per se, but not always with satisfactory results; such works are commented upon, with corrections offered, where relevant. The present book therefore seeks, among other things, to bring correct versions of the facts about the Balkan languages to the general linguistic public by way of setting the record straight but also making available in as accurate a manner as possible the interesting facts from the Balkans that can inform and enlighten more general matters in linguistics.

The relevance of linguistic contact for the history of specific languages is often elided or minimalized in the linguistic equivalents of national histories (cf. K. Brown 2015).[12] Such works attempt to ignore or explain away the possibility that multilingualism could be responsible for change in their respective national languages (cf., e.g., Kazazis 1977 on Greek denials of Turkisms in Greek). This approach is epitomized in Andriotis & Kourmoulis' 1968: 30 conclusion to their plenary paper at the first international Balkan Congress in 1966:

> Notre opinion personelle sur ce problème capital est que l'unité linguistique de nos peuples est une fiction qui n'est perceptible que de très loin; en isolant les traits comuns, que nous avons examinés, du corps de chaque langue, on a l'impression illusoire que les similitudes l'emportent sur les differences. Mais dès qu'on s'approche de la structure spécifique de chacune de ces langues on voit qu'en réalité les similitudes sont relativement très peu nombreuses, elles ne résident pas dans l'intérieur de la langue, elles sont tout à fait inorganiques et superficielles en face des différences et des oppositions, qui sont au contraire innombrables et profondes, et que les traits communs, quelque soit leur nombre, s'effacent devant l'immensité des éléments particuliers, dus à la difference d'origine et d'histoire de chaque langue.
>
> Our personal opinion on this major problem is that the linguistic unity of our peoples is a fiction that is only visible from a great distance; by isolating those traits which we have examined from the body of each language, one has the illusory impression that the similarities prevail over the differences. But as soon as one approaches the specific structures of each of these languages one sees that the

11 Thus for example, Comrie 1976: 108–110 cites Bulgarian in a discussion of perfect aspect and inferentiality – a welcome citation to be sure – but in his presentation of evidentiality in Bulgarian he follows normative grammar and thus elides the problems presented by actually occurring data; see Aronson 1967, 1991; Friedman 1982a, 2002b; Alexander 2000c, 2002; Fielder 2002; McClain 2002.

12 We note, for instance, that two otherwise deeply scholarly treatments of the historical development of Modern Greek (Tonnet 1993 and Horrocks 2010) pay little or no attention to possible structural effects on Greek of contact with other, co-territorial Balkan languages (as discussed in Joseph 1994c on Tonnet and in Joseph 2001f, reviewing Horrocks 1997, the prior first edition to Horrocks 2010), although there is no overt hint of anything like the nationalistic ideology that characterizes some other approaches.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

> similarities are relatively much less numerous. They do not reside in the interior of the language; they are entirely inorganic and superficial when compared to the differences and oppositions which, on the contrary, are innumerable and profound. These similarities, whatever their number, disappear before the immense number of specific elements that are due to the differences in origin and history of each language.

Such ideologically determined approaches to language contact (cf. Friedman 1997a) are also connected with other – often interrelated – issues of linguistic social history such as language planning (e.g., Friedman 1989a, 1993b), language shift (e.g., Tsitsipis 1998), linguistic purism (e.g., Friedman 1989b; Heyd 1954; Lewis 2000), and linguistic nationalism (e.g., Friedman 1986b; Kazandžiev 1943; Joseph 1985a). These are issues that typically have not been addressed in the usual handbooks on Balkan linguistics but are of enormous relevance to a thorough account of the field.

Similarly, more detailed discussion of relevant languages other than those that are routinely treated in the standard handbooks, such as Romani, Judezmo, and Balkan Turkic, as noted above, is sorely needed.[13] Of the various handbooks cited above, only Desnickaja & Tolstoj 1998 treats Judezmo and only Hinrichs 1999a treats Romani and Balkan Turkish (but not Gagauz), but neither of these integrates these languages into a unified Balkan linguistic account.[14] Moreover, in the period since the emergence of the modern Balkan standard languages, their relation to dialectology and to competition among diasystems (cf. Ivić 1962) has only rarely been addressed. The overwhelming majority of publications in Balkan linguistics in the past fifty years have been based almost exclusively on standard languages (Sawicka 1997, 2014, 2023 are notable exceptions to the general trend). As Heřman 1968: 342 observes, however, "[a] historical study of Balkan convergences is possible only on the basis of both diachronic and synchronic dialectology." We rectify this situation by integrating the under-studied and under-described languages of the sprachbund into our account, and by giving consideration to regional and other (e.g., social) dialects as needed.

After seven decades of calls for a Balkan linguistic atlas (Anagnostopoulos 1935: 273; Małecki 1935: 76; Deanović 1961; Hamp 1989a; Friedman 2000d), an international group of Balkanists attempted finally to address the task through a pilot project begun at the University of St. Petersburg in 1996, the *Malyj dialektologičeskij atlas balkanskix jazykov* 'Small dialectological atlas of the Balkan languages' (Domoselickaja 1997; Sobolev 1997, 1998b, 1999; Plotnikova 1996; Rusakov & Sobolev 2003). This project, however, pursued the more modest goal of working to produce an atlas that would cover selected, representative

13 We should note here that, as in the case of Posner 1996 for Romance, the goal of our book is not a set of complete grammatical descriptions. For these we refer readers to the references given in Friedman 2013a.

14 Asenova 2002: 220 mentions the Balkan type of analytic future found in the Balkan and South Vlax dialects of Romani, but she does so in the context of excluding Romani from the inventory of Balkan languages (see Friedman 2006d for a response). Although Tomić 2006 includes Romani, her data are flawed by numerous errors.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

dialects in depth, rather than attempting the Herculean task of surveying all the dialects of all the languages of the region. For this purpose, a total of twelve points were chosen, seven Slavic points (three in Bulgaria, one each in Croatia, Montenegro, and Serbia, and North Macedonia), two Albanian, two Greek, and one Aromanian (in Greece).[15] Moreover, since the focus of the project is precisely areal, it examines the dialects in terms of those categories that are of greatest relevance to language contact: syntax and lexicon (see Sobolev 2001a). We can also mention here the Romani Morpho-Syntax Database at the University of Manchester (Matras & Elšík 2001–2005, 2008, hereafter and listed in the references under RMS) and the Romani Lexical Database at the University of Graz (Halwachs et al. 2006, hereafter and listed in the references under ROMLEX), which offer coverage in these same areas for Romani dialects.[16] Material from these projects also informs our work.

In addition to examining neglected speech forms, we also devote attention to subdisciplinary approaches that have not received adequate treatment in a Balkan context. For instance the study of the role of pragmatics as the entry point for structural change (cf. Mithun 1992; Prince 1988; Friedman 1994a) as well as more modern approaches to syntax are other areas addressed here for the first time in a general Balkan linguistic handbook.

Finally, the social setting of the Balkan languages and the relationship of sociolinguistic issues in general to the field of Balkan linguistics have only relatively recently begun to be explicitly addressed in the handbooks (Hinrichs

15 The villages in question (listed in the form given in the series itself and identified (in parentheses) by dialect grouping and region) are the following: Slavic – Otok (Neo-Štokavian Ikavian, Sinj district, Dalmatia, northwestern Croatia), Kamenica (Timok, Knjaževac district, eastern Serbia), Zavala (Zeta-Lovćen, Piper tribe, Podgorica district, southeastern Montenegro), Ravna (Moesian, Provadija region, northeastern Bulgaria), Gela (with Široka Lăka; central Rhodopian, Smoljan district, southern Bulgaria), Gega (Pirin Macedonian, Petrič district, southwestern Bulgaria), Peštani, (western Macedonian, Ohrid district, southwestern Republic of North Macedonia); Albanian – Fushë Muhurr (henceforth Muhurr; West Central Geg, Peshkopi district, west central Albania), Leshnja (northern Tosk, Skrapar district, southeastern Albania); Greek – Erátyra (Mac Selce or Selsko; northern Greek, Kozani district, Aegean Macedonia, northwestern Greece); Kastélli (southern Greek, northern Peloponnese, southern Greece); and Aromanian – Turia (Grk Krania or Kranéa; Pindus group, southern Aromanian, Grevena (Aro Grebena) district, Aegean Macedonia, northwestern Greece). In addition to the four villages that are specified as being in Macedonia *sensu lato*, Muhurr, on the west bank of the Black Drin, was in the Ohrid *Kaza* for part of the nineteenth century. The atlas thus correctly concentrates on villages in the heartland of Balkan contact, although the villages themselves are monoglot, except for Turia, where Greek is now in the process of replacing Aromanian (Bara et al. 2005: 18). There are also a number of comparative volumes in the series dealing with lexicon and culture (Sobolev 2005–2013; Domoselickaja 2012; and Domoselickaja & Sobolev 2018) and the noun (Sobolev 2005c).

16 It is worth noting that syntax and morphosyntax are precisely the areas least covered by traditional dialect atlases, which focus on phonology and morphophonology, with some data on morphology and lexicon. See the references in Friedman 2013a, to which can be added Saramandu & Nevaci 2014a and Atanasov 2008, 2013, 2015. Here we can also note that Risteski 2019, 2021 also provide a useful model for ethnolinguistic dialect atlases. Additional volumes of Macedonian ethnolinguistic dialect atlases have been and are being produced by Risteski, and as of this writing Evangelia Adamou and Andrey Sobolev are editing an *Atlas of the Balkan Linguistic Area* (ABLA), to be published by De Gruyter and available online, covering selected features for almost sixty points of various Balkan languages and also dialects spoken outside the Balkans.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002

1999a), although Morozova & Rusakov 2018a, 2018b; Sobolev et al. 2018 and the articles in Sobolev 2021c are all relevant here. The present work is therefore more than a description of the structural convergences and divergences of the Balkan languages; it is also a history of the social implications of both changes and conservatisms.

We thus invite the reader to use this study as a vehicle for achieving the goal set out in the epigram at the head of this section, but also for entering into this fascinating – and instructive – part of the (linguistic) world.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 10:29:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.002