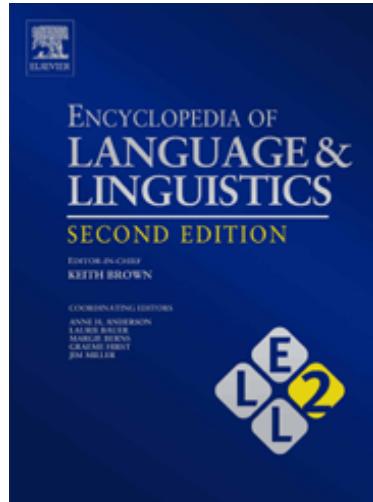


Provided for non-commercial research and educational use only.  
Not for reproduction or distribution or commercial use



This article was originally published in the *Encyclopedia of Language & Linguistics, Second Edition*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

Friedman V A (2006), Balkans as a Linguistic Area. In: Keith Brown, (Editor-in-Chief) *Encyclopedia of Language & Linguistics, Second Edition*, volume 1, pp. 657-672. Oxford: Elsevier.

- Hunter T M (1988). Balinese language: historical background and contemporary state. Ph.D. diss., Michigan University.
- Kersten J (1970). *Tatabahasa Bali*. Flores: Arnoldus.
- Kersten J (1984). *Bahasa Bali*. Ende-Flores: Nusa Indah.
- Mbete A M (1990). Rekonstruksi protobahasa Bali-Sasak-Sumbawa. Ph.D. diss., Universitas Indonesia.
- Oka Granoka I W, Udara Naryana I B & Jendra I W (1985). *Tata bahasa Bali*. Denpasar: Proyek Pengembangan Bahasa dan Sastra Indonesia dan Daerah Departemen Pendidikan dan Kebudayaan.
- Pastika I W (1999). Voice selection in Balinese narrative discourse. Ph.D. diss., Australian National University.
- Shadeg N (1977). *A basic Balinese dictionary*. Denpasar: Dharma Bhakti.
- Simpson I W (1985). *Kamus Bahasa Bali*. Denpasar: PT Mabharti.
- Sutjaja I G M (2000). *Practical Balinese-English-English-Balinese dictionary*. Denpasar: Bali Post.
- Sutjaja I G M (2003). *Kamus sinonim bahasa Bali*. Denpasar: Lotus Widya Suaru/Universitas Udayana.
- Sutjaja I G M (2004). *Kamus Indonesia-Bali-Inggris*. Denpasar: Lotus Widya Suaru/Universitas Udayana.
- Van Der Tuuk H N (1897). *Kawi-Balineesch-Nederlandsch Woordenboek*. Batavia: Landsdrukkerij.
- Warna I W et al. (1993). *Kamus Bali-Indonesia*. Denpasar: Dinas Pendidikan Dasar Propinsi Dati I Bali.
- Wechsler S (1999). 'HPSG, GB, and the Balinese Bind.' In Kathol A, Koenig J-P & Webelhuth G (eds.) *Lexical and constructional aspects of linguistic explanation*. Stanford: CSLI. 179–195.
- Wechsler S & Arka I W (1998). 'Syntactic ergativity in Balinese: an argument structure based theory.' *Natural Language and Linguistic Theory* 16, 387–441.

## Balkans as a Linguistic Area

V A Friedman, University of Chicago, Chicago, IL, USA

© 2006 Elsevier Ltd. All rights reserved.

### Definitions

#### Sprachbund

Among the proposed glosses for *sprachbund* are 'linguistic league', 'linguistic area', 'convergence area', and 'diffusion area', but here I will treat *sprachbund* as a loanword into English, like the French *genre*, so henceforth it will be neither capitalized nor italicized. In modern terms, a sprachbund is understood as two or more geographically contiguous and genealogically different languages sharing grammatical and lexical developments that result from language contact rather than a common ancestral source. (Some linguists set the minimum number at three, but I would argue that the convergent and diffusion processes constitutive of a sprachbund are the same for two languages as for three.) In his original formulation of the concept, first in 1923 in a Russian journal article and again in 1928 at the first International Congress of Linguists, N. S. Trubetzkoy used Bulgarian as his example of a language that belongs to the Slavic linguistic family and at the same time to the Balkan sprachbund. In the case of the Balkan sprachbund, the languages are in fact all Indo-European (excluding Balkan Turkic), but they belong to groups that were separated for millennia, and thus, upon coming back into contact, had become sufficiently distinct for contact phenomena to be distinguished from inherited phenomena.

#### Balkan

The use of the term 'Balkan' (from Turkish, *balkan* 'forested mountain', also the name of a mountain range in Central Bulgaria) to refer to the peninsula also known as Southeastern Europe dates from the 19th century, when European attention turned to Ottoman Turkey, which then included most of what became the Balkan states. As a geographic entity, the Balkan peninsula is unproblematically defined on three sides as the land mass defined by the Adriatic, Mediterranean, and Black Seas, but the northern geographic boundary cannot be set in any nonarbitrary way that is applicable without qualifications in terms of either politics or linguistics. In modern geopolitical terms, from the 1920s to 1991, the Balkans were most frequently understood as comprising Albania, Bulgaria, Greece, Romania, Turkey in Europe, and former Yugoslavia.

#### The Balkan Languages

For linguistics, the Balkan sprachbund has traditionally consisted of Albanian, Greek, Balkan Romance (BR), and Balkan Slavic (BS). Albanian is divided into two dialects, Gheg north of the river Shkumbi and Tosk south of it. The modern standard is based on northern Tosk. Mainland Greek is also divided between northern and southern dialects at the Gulf of Corinth and the northern frontier of Attica, the southern dialects of the Peloponnese being the basis of the standard vernacular Dhimitiki. During the 19th century, Modern Greek was still called *Romaic*, i.e., 'Roman', a reference to Byzantium as the second

Rome. BR consists of Romanian, Aromanian, Megleno-Romanian (MR), and Istro-Romanian. Dalmatian, a remnant of West Balkan Romance, whose last speaker died in 1898, is rather poorly attested and generally does not figure in Balkan linguistic accounts. Istro-Romanian is, like Arbëresh (the Albanian of Italy) and Asia Minor Greek (until the exchange of populations between Greece and Turkey in 1923), outside the Balkan geolinguistic area (see 'Balkan Languages vs. Languages of the Balkans'). The Romanian standard is based on the Wallachian dialects of the south, as is the standard of the Republic of Moldova, which at various times has called its official language Moldovan or Romanian. (At present [31 October 2004] the official name is Moldovan.) Aromanian, spoken in Albania, Greece, the Republic of Macedonia, and southwestern Bulgaria (with a large diaspora in Romania, especially Dobrogea) is divided into north/west dialects of Albania and western Macedonia and south/east dialects of Greece and eastern Macedonia. A standard based primarily on the eastern dialect is in use in the Republic of Macedonia. MR survives in seven villages near Gevgelija in the southeast of the Republic of Macedonia and across the border in Greece. During the 19th century, BR was often called Wallachian. The term 'Vlah' can be used as a convenient cover term for BR south of the Danube (Aromanian plus Megleno-Romanian). BS consists of Bulgarian, Macedonian, and the southeast Serbian (Torlak) dialects. Bosnian/Croatian/Serbian (BCS) together with Slovene, form the West South Slavic group, and Macedonian and Bulgarian comprise East South Slavic. The Bulgarian standard is based on its eastern dialects, the Macedonian standard on its west-central dialects. The northern and western boundaries of Torlak as a Balkan dialect are variously defined using phonological or morphological criteria. The narrowest definition is morphological, e.g., the isogloss for the presence of the postposed definite article; the broadest definition is phonological, e.g., the absence of distinctive vocalic length and tone. During the 19th century, BS was often called 'Bulgarian,' and Bulgarian and Serbian linguists and armies fought over where to draw a line between Bulgarian and Serbian. Unable to adjust to modern times, many Bulgarian linguists still cling to the 19th-century practice.

**Romani** Despite having been summarily dismissed by traditional Balkan linguists, Romani in the Balkans displays many of the same contact-induced structural phenomena and is increasingly present in Balkanological works. Two of the four main dialectal groups of Romani are spoken in the Balkans: Balkan and Vlax (not to be confused with Vlah). The Vlax

dialects of Romani take their name from the fact that they took shape in Romania, but they are now dispersed all over Europe and beyond. In the Republic of Macedonia, a Romani standard is emerging on the basis of the Arli dialect of the Balkan group. Unless otherwise specified, references to Romani refer to those dialects spoken in the Balkans.

**Turkish** Balkan Turkish is divided into two major dialect groups: West Rumelian Turkish (WRT) and East Rumelian. The boundary between the two corresponds roughly to the east-west line of Bulgarian dialects. The Christian Gagauz of Bulgarian and Romanian Dobrudja and Gagauz Yeri in Moldova and adjacent parts of Ukraine speak a language in the Oghuz group – to which Turkish also belongs – which was recognized as official in the USSR in 1957. Although most Balkan linguistic studies treat Turkish as an adstratum, contributing lexicon and phraseology but very little else (aside from evidentiality, see 'Evidential' below), WRT and Gagauz also participate to a certain extent in the Balkan sprachbund. Most of Gagauz, however, ended up in the former Russian Empire, due to migration and border changes. As a result, most of Gagauz is now more influenced by Russian, while the dialectal Gagauz remaining in the Balkans is in need of description.

**Jewish Languages** Judezmo, the language of the Jews expelled from Spain in 1492, became the majority language among Balkan Jews, overwhelming Judeo-Greek (Yavanic, Yevanic), which survived in the Romaniote liturgy and some enclaves in Epirus. (A written version of Judezmo based on literal translation from Hebrew is known among scholars as Ladino.) Although most speakers of both Judezmo and Judeo-Greek were murdered in the Holocaust, these languages survive as endangered languages and also participated in Balkan linguistic processes.

**Balkan Languages vs. Languages of the Balkans** There are many other languages spoken in the Balkans in enclaves with varying social relations, e.g., Armenian, Circassian (until 1999), German, Hungarian, Ruthenian, Tatar, Ukrainian, Yiddish, etc. Aside from the dialects spoken in Romania, most of these are outside the geolinguistic Balkans, which for our purposes has a northwest boundary defined by contiguous Albanian dialects that join the major Torlak isoglosses continuing to the Danube. (Such a definition includes the southernmost Slavic dialects of Montenegro as well as the Slavic dialects of northern Kosovo, neither of which fall in the Torlak group. In terms of the Balkan sprachbund, these dialects do show some important transitional features, which will be

noted.) For the most part, the enclave languages were late arrivals or outside the area of intensive diffusion/convergence and did not participate in the type of complex Balkan multilingualism that characterizes the sprachbund as a whole. We can thus distinguish Balkan languages, i.e., those in the sprachbund, from languages of the Balkans, i.e., languages spoken in the Balkan peninsula.

## History of Balkan Linguistics

### 1770–1861

The earliest collections of Balkan linguistic material were intended to eliminate Balkan linguistic diversity. The 1770 Greek-Aromanian-Albanian vocabulary of T. Kavaliotis and the 1793 or 1794(?) Greek-Aromanian-Macedonian-Albanian lexicon of Daniil of Moschopolis (Albanian Voskopoja) were explicitly aimed at the Hellenization of the speakers of other Balkan languages. The first was republished in 1774 by J. Thunmann, who was the first to suggest that Albanians and Romanians were descended from Illyrians, Dacians, and Thracians, thus laying the groundwork for the substratum theory of Balkan linguistics. The second was republished in 1814 by M. Leake, who suggested that similarities among Albanian, BR, and Greek were due to BS influence. His one concrete example was the postposed definite article. It was this same phenomenon that most impressed J. Kopitar, whose 1829 characterization of BR, BS, and Albanian as *drey lexikalisch verschiedenen, aber grammatisch identischen Sprachen* ‘three lexically distinct but grammatically identical languages’ – which he attributed to the influence of a Thraco-Illyrian substratum – is taken as the earliest formulation characterizing the Balkan sprachbund. Kopitar also noted the replacement of infinitival with subjunctive constructions and the formation of the future using ‘want’ as shared with Greek and Serbian as well.

A. Schleicher is sometimes cited as the first to formulate the Balkan sprachbund in 1850, when he writes of Albanian, BR, and BS saying *eine Gruppe aneinandergränzender Sprachen zusammengefunden hat, die bei stammhafter Verschiedenheit nur darin übereinstimmen, dass sie die verborbensten ihrer Familie sind* (‘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’). However, since he gives no indication of the causes of this ‘corruption’, his formulation differs from Kopitar’s mainly in its ideology of language change as degeneration.

The next real advance in the development of Balkan linguistics was F. Miklosich’s 1861 article on Slavic elements in Romanian, which added genitive-dative merger (see ‘Genitive-Dative Merger’), object pronoun doubling (see ‘Resumptive Clitic Pronouns [Reduplication, Replication]’), and the formation of teens (see ‘Numeral Formation: The Teens’). Miklosich accorded more attention to Greek and was also the first to adduce a number of phonological changes, including the development of stressed schwa (see ‘Vowel Reduction and Raising’) and the raising of unstressed /a/ and /o/ to schwa and /u/, respectively (see ‘Stressed Schwa’).

### 1861 Onward

The next six decades were characterized by the gathering of materials relating to specific Balkan languages or specific aspects of individual or pairs of Balkan languages. The 1920s saw the basic syntheses and theoretical formulations that continue to inform the field. Trubetzkoy’s contribution has already been described. In 1925, A. Seliščev attempted a balanced account of Turkish, Slavic, Latin, Greek, and substratum languages as the sources of various Balkanisms, i.e., the similarities among the Balkan languages that can be attributed, at least in part, to shared, contact-induced change. Sandfeld (1930) tried to attribute almost all the commonalities of the Balkan sprachbund to the influence and prestige of Byzantine Greek. Other scholars have laid particular emphasis on Balkan Latin as the primary causal factor, while our knowledge of the pre-Latin non-Hellenic languages of the Balkans remains too meager for almost any serious speculations beyond the lexicon.

While the 1920s saw the establishment of Balkan linguistics as a subdiscipline within linguistics, the period from 1930 to 1960 was characterized by slow growth and was also the period when the insights gained in Europe finally came to the attention of North American linguists. From the 1960s onward, there has been a constant increase in the production of studies pertaining to the Balkan languages and Balkan linguistics. At the same time, studies of such contact-induced phenomena as creolization, code switching, and language shift have led to the identification of contact linguistics as an overarching field of study. More recently, in the past decade or so, a renewed interest in linguistic typology has brought forward questions of the extent to which the Balkan sprachbund is or is not part of a larger European linguistic area, defined more by typological profile without necessarily identifying specific paths of diffusion or convergence. We will return to the question of Eurology vs. Balkanology in ‘Causation’.

## Balkanisms

This section surveys some of the principal Balkanisms (see ‘1770–1861’) as identified during the course of the past two centuries. Although system, not mere inventory, must be the basis of detailed study, and a given surface phenomenon may function differently in different systems, it is nonetheless convenient to use lists as a kind of shorthand for the systemic relations that can yield the most insights. We do not want to fetishize the labels for these systemic manifestations, assigning numeric values to them and tallying up the number of points a language ‘scores’. Rather these labels stand for complex interrelations that include differences as well as similarities that must be elucidated in their larger contexts (cf. Friedman in Reiter, 1983).

## Phonology

In contradistinction to linguistic areas such as the Caucasus, the Northwest Coast, and South Asia, where phonological features such as glottalization and retroflexion are among the most salient commonalities, there are no truly pan-Balkan phonological features. Rather, there are articulatory tendencies of greater or lesser extent.

**Vowel Reduction and Raising** The reduction of unstressed vowels to schwa or nonsyllabic elements (and thence sometimes to zero) as well as the raising of unstressed mid-vowels (/e/ and /o/) to high vowels (/i/ and /u/, respectively) can be treated as Balkan, albeit not pan-Balkan. Both Albanian and BR show a tendency to reduce unstressed vowels as early as the Latin period, e.g., Lat. *imperātor* > Albanian *mbret* and Romanian *împărat* ‘king’. While shared phonological tendencies in Albanian and BR, like shared vocabulary of pre-Latin origin, are attributed by some scholars to substrate influence, the evidence of vowel reduction in Western Romance leads other scholars to suggest that this is a typological rather than an areal feature. Nonetheless, the raising and/or elimination of unstressed vowels is characteristic of southeastern Macedonian, eastern Bulgarian, northern Greek, BR, and Gheg, although the details differ among these languages.

**Stressed Schwa** All the Balkan languages and their dialects possess the classic European five vowel system /a, e, i, o, u/, at least under stress. A phenomenon common in the Balkans is the existence of a stressed schwa, but its status as a contact-induced phenomenon is not pan-Balkan. Greek lacks stressed schwa altogether. In Macedonian, almost all the dialects outside the west-central area have stressed schwa, but of different origins in different areas, and some

western peripheral dialects also lack stressed schwa. Most of Bulgarian has stressed schwa, but not the Teteven-Erkech and central Rhodopian dialects. In Albanian, stressed schwa develops from nasal *â* only in Tosk, but it is incorrect to characterize all of Gheg as lacking stressed schwa, since it also occurs in central Gheg as a result of later processes. Romani has schwa when in contact with languages that have it. WRT has a tendency to lower and front the high back unrounded vowel to schwa.

**Other Vowels** Most Balkan languages lack front rounded vowels, but most of Albanian has /ü/, or, in West Central Gheg, /ö/. Southern Montenegrin dialects in contact with Albanian also have /ü/, but East Central Gheg, which is mostly in Macedonia, unrounds /ü/ to /i/, as does southernmost Tosk (Lab, Çam, Arvanitika), in contact with Aromanian and Greek (which also merged /ü/ with /i/, a change that had not yet been completed in the 10th century). Similarly, WRT tends to eliminate /ö/ by merging it with /o/ or /ü/ (more rarely /e/), and /ü/ (like /u/ and /i/) becomes /i/ word finally. Other vocalic phenomena that have been suggested are relatively localized.

**Consonants** The alternation of clear /l/ before front vowels and velar /ɫ/ elsewhere is characteristic of BS (including Torlak but not the rest of BCS), Northern Greek, Balkan Romani, and Vlah, but not Albanian, where the two sounds are in phonemic contrast, nor Daco-Romanian and Southern Greek, where only clear /l/ occurs. Aromanian has Greek and Albanian interdental and Greek voiced velar and palatal fricatives in loanwords from Albanian and Greek, but these tend to be replaced by corresponding stops and the palatal glide by speakers who do not know Greek or Albanian, particularly the younger generation in Macedonia.

Aside from Greek, most Balkan languages have an opposition between strident palatal affricates, on the one hand, and mellow palatals, dorso-palatals, or palatalized velars, on the other. The opposition is neutralized in Albanian, BS, and WRT dialects in Kosovo, parts of Western Macedonia, and along the Serbo-Bulgarian border. Northern Greek has palatals lacking in the south.

In western Macedonia, the velar fricative is generally lost or replaced in Albanian, Macedonian, and WRT, a phenomenon that extends into parts of Kosovo, as well as adjacent Serbia, much of Montenegro, and Bosnia-Herzegovina, where the preservation of BCS /x/ is characteristic of Muslim and some Catholic dialects now Bosnian and Croatian, respectively.

In the northern Gheg of Malësia e Madhe, final devoicing is a phenomenon shared with adjacent Montenegrin dialects. It is worth noting that final devoicing is atypical for most of the rest of BCS and Gheg, and it appears rather to be a Macedonian feature extending into this region. Such influence also seems to be the case in the transitional Gheg and northern Tosk dialects. Some of the Romani dialects in this region also have final devoicing, and in the WRT of these regions final devoicing, which is usually limited to stops in Turkish, extends to fricatives. Five of the seven MR villages also have final devoicing.

**Prosody** Although prosodic distinctions of length, and in some cases pitch, were present in the attested ancestors of the Balkan languages, the modern Balkan languages are generally characterized by the absence of length and tone and the presence of a stress accent that usually does not move further back in the word than the antepenultimate syllable. If stress does move further back, there is usually a secondary stress on one of the last three syllables. However, Northern Gheg and Southern Tosk preserve Common Albanian length, and Southeastern Macedonian has new long vowels as the result of loss of intervocalic consonants and elision. Similar new long vowels occur in Gora, a string of Slavic-speaking Muslim villages along the western and northern slopes of Mounts Korab and Šar in northeastern Albania and the southwestern corner of Kosovo. The most significant isoglosses (fixed antepenultimate stress, postposed article, etc.) link Goran with the northwest Macedonian dialects rather than with the Serbian of Prizren.

### Morphosyntax

**Grammaticalized Definiteness** In BS, BR, and Albanian, native demonstrative pronouns have been encliticized or suffixed to nominals (normally the first in the noun phrase) and become definite articles. The article follows a plural marker, if any, and in BS the clitic-like nature of the article is seen in that it does not trigger certain morphophonemic alternations, e.g., Macedonian *starec* ‘old man’, *starci* ‘old men’ but *starecot* ‘the old man’ and not \**starcot*. Hamp (1982) adduces evidence suggesting that the autochthonous language that became Latinized into Romanian and with which the ancestor of Albanian was in contact might already have had a postposed definite article by the time of contact with Latin. Common Slavic already had a postposed relative pronoun \**jí* affixed to adjectivals to denote definiteness, as this phenomenon is attested in Old Church Slavonic (OCS; 9th–11th centuries), and the morphology (but not the grammatical meaning) survives in Slavic outside the geopolitical Balkans. Remnants of this

older definite/indefinite opposition survive in West South Slavic adjectives, and traces of the morphology occur in BS, e.g., Macedonian *star* ‘old INDEF.MASC’, *stariot* ‘old DEF.MASC’, where the /i/ indicates that the newer definite article has been suffixed to a definite adjectival form. Scandinavian and dialectal North Russian also have postposed definite articles of pronominal origin, and Czech, which has been in close contact with German, has uses of its deictics that are basically articular. These typological parallels and historical antecedents, however, do not change the fact that the BS postposed definite article developed during the period of its contact with BR and Albanian.

Greek and Romani have preposed definite articles, both based on native material. In the case of Greek, the pronoun that became an article was still mostly demonstrative and was facultative except with proper names in Homeric, but it was obligatory in Attic. Romani articles look like borrowings from Greek, e.g., MASC NOM SG o FEM NOM SG *i*, but the oblique forms /le/ and /la/ in Vlax dialects demonstrate that the Romani articles are derived from native demonstratives, reflecting the regular change of \**t > l*, which occurred prior to contact with Greek. It was contact with Greek, however, that triggered the transformation of native material into definite articles, and Romani usage patterns very much like Greek. Romani dialects outside the Balkans in contact with languages lacking definite articles tend to lose them.

The use of an atonic form of the numeral ‘one’ as an indefinite article is characteristic of the Balkan languages and, even though such developments are common in many languages, is arguably a Balkanism. ‘One’ was not used in this function in OCS, Ancient Greek, or Latin, but it was so used in Orkhon Turkic (8th century C.E.). To this we can add the fact that such usage does not occur in East Slavic. Usage in Turkish, Albanian, and BR is at a similar level of frequency to that of English, although details in individual grammars will cause some lack of isomorphism. Usage in BS and in Greek is approximately half that of the other Balkan languages, while usage in Romani in the Balkans patterns with BS and Greek, and Romani elsewhere patterns like its contact languages. An indication that this is an areal phenomenon despite the occurrence of such usages in Western Europe and elsewhere is the fact that, as one moves north and east through West South Slavic territory, the usage becomes increasingly restricted.

Finally, we can mention here the phenomenon of double determination, i.e., the presence of a definite article on a noun modified by a demonstrative pronoun. Such usage occurs in Greek, BR, BS, Albanian,

and Romani, although the rules and relative frequency and acceptability of the construction vary. In Greek it is obligatory, e.g., *autós o ánthrōpos* or *o ánthrōpos autós* but not \**autós ánthrōpos* ‘this person’. In Romanian, the article is not used if the deictic is preposed, but is used if it is postposed (and the deictic takes the so-called deictic particle *-a*): *omul acesta* but *acest om* ‘this person’, cf. Aromanian *aistă carte, cartea aistă* ‘this book’. Megleno-Romanian has frequent double determination *tsista lup-u* ‘this wolf-DEF’, but indefinite nouns also occur *tsista drāc* ‘this devil-INDEF’. In Albanian, the deictic is preposed to either the indefinite or definite: *ai njeri, ai njeriu* ‘this person’. In BS, double determination occurs but is considered dialectal, Macedonian *ovoj čovekov* (vs. *ovoj čovek*) ‘this person’, or Torlak *taja starata* ‘that old [lady]’. Romani permits but does not require the use of a definite article with a demonstrative, in which case the article must precede the substantive but the demonstrative can precede or follow: *kova manuš, kova o manuš, o manuš kova* ‘this person’. Double determination or the order noun-determiner is pragmatically more thematic in the discourse.

**Resumptive Clitic Pronouns (Reduplication, Replication)** Balkan languages are characterized by the use of clitic or weak resumptive object pronouns that agree in gender, number, and case with the non-clitic/strong pronoun or substantive they refer to. This phenomenon is called (object/pronoun) reduplication/doubling in Balkan linguistics and is connected to expressions of definiteness, referentiality, and animacy: the first candidates for reduplication are personal pronouns (inherently definite and, in the first two persons, usually human), then indirect objects (usually human, often topicalized), then definite direct objects, and finally specific or topicalized direct objects.

From a morphosyntactic point of view, there are four types of reduplication: pronominal object doubling, substantival object replication, pronominal possessive doubling, and substantival possessive replication. All four phenomena can be illustrated in the following Macedonian sentence:

Tatko mi	moj	i	majka
<i>father me.DAT</i>	<i>my.M</i>	<i>and</i>	<i>mother</i>
mu	na car-ot	im	rekoa
<i>bim.DAT</i>	<i>to king-the</i>	<i>them.DAT</i>	<i>said.3PL.AOR</i>
nim	da	mu	gi
<i>them.DAT</i>	<i>SP</i>	<i>him.DAT</i>	<i>them.ACC</i>
dadat	knigi-te	na	dete-to
<i>give.3PL.PRES</i>	<i>books-the</i>	<i>to</i>	<i>child-the</i>
'My father and the king's mother told them to give the books to the child.'			

The first three of these expressions are facultative and could be replaced by *tatko mi, majkata na carot* (*majka* is definite), and *im*, respectively. The reduplication serves to emphasize or focus the referent of the reduplicated pronoun. The last set of reduplications, *mu ... na deteto* and *gi ... knigite*, are obligatory in standard Macedonian and, for the most part, in the western dialects on which it is based. The norm requires reduplication for definite direct objects and all indirect objects. In practice, however, even the most normative grammar shows that specificity or topicalization rather than definiteness is the trigger (Koneski, 1967: 232):

kako vistinski ja doživuvame edna situacija  
how truly it.ACC experience one situation  
1PL.PRES  
'how we actually experience a [given] situation'

Pronominal object doubling occurs in all of BS (and southern Montenegro), BR, Albanian, Greek, and Romani. It is conditioned by discourse factors such as emphasis or focus and can be compared to the use of subject pronouns. Just as the fact that the subject is marked on the verb makes the subject pronoun redundant unless there is a need for emphasis or specification, so, too, the clitic pronominal object, which is the required form if the object is a pronoun, makes the full form redundant except under similar discourse-bound circumstances. The absence of such doubling from the rest of BCS is a diagnostic separating Balkan from non-Balkan Slavic.

The clitic replication of oblique nominals shows how grammatical change can enter a language via discourse phenomena and at the same time supports Topolińska's observation that analytic markers of referentiality are characteristic of convergent development. Object reduplication is another scalar Balkanism. It is rare in Torlak and used only for emphasis and thus separates East from West South Slavic. Similar conditions hold for Romani except in possessive constructions. Object reduplication is more pragmatically conditioned and less grammaticalized in Bulgarian, Romanian, and Greek, where the phenomenon signals topicalization, focus, or emphasis, and is restricted by factors such as animacy (or humanness) and degree of referentiality (definiteness, specificity, determinacy, etc.). In Albanian, Vlah, and West Macedonian, reduplication has become grammaticalized. It is most frequent in Macedonian, where, unlike in the other Balkan languages, it can even occur ( facultatively) with indefinite indeterminate pronouns such as *nikoj* 'nobody'.

While it lacks a definite article, Turkish does have a special accusative marker used for definite or specified direct objects. The following proverb illustrates

how the Turkish definite accusative is rendered by Balkan object reduplication. Note that Greek and Bulgarian have reduplication with an indefinite object, indicating its specificity:

Turkish:	Yavaş başı kilç kes-mez (Turkish) gentle head- sword cuts-not DEFACC
Bulgarian:	Pokorena glava sabja ne ja seče bent head sword not it.ACC cuts
Greek:	Kefáli proskynéméno spathì dhèn head bent sword not tò kóvei it.ACC cuts
Romanian:	Cap-ul plecat nu l taie sabia head- bent not it.ACC cuts sword. DEF DEF
Albanian:	Kokën e falur yatagan-i head PART.FDEFACC bent sword-DEF nuk e pret not it.ACC cuts 'A/The sword does not cut off a/the bent head' (= Keep your head down.)

Possessive doubling is a more restricted phenomenon. The use of dative clitics to indicate possession in Macedonian is limited to kinship terms, Aromanian has special possessive clitics that can only be used with kinship terms, and Albanian also has special possessive constructions for kinship terms. In Bulgarian, possession is usually signaled by a dative clitic following the definite form of the noun, and possessive adjectives, which are the norm in Macedonian, are more emphatic in Bulgarian. In Greek, clitic dative pronouns after the definite form of the noun is the normal manner of indicating possession, and emphasis is rendered by adding the appropriate form of the adjective *dikós* '[one's] own' immediately before the pronoun. However, pronominal doubling is also used colloquially for emphasis:

to vivlio mou mena
the book me.GEN me.GEN
'my book'

Romanian also has such clitic doubling colloquially:

propria-mi mea semnătura
own.FEM-me.DAT my signature.DEF
'my very own signature'

Substantival possessive replication occurs in all the Balkan languages, but the details differ from language to language. The Turkish construction of genitive possessed plus pronominal suffix on the possessor is the normal pattern:

kral-in anne-si
king-GEN mother-his
'the king's mother = the mother of the king'

**Genitive-Dative Merger** Albanian, BS, BR, and Greek have no formal (i.e., surface) distinction between the shape of the genitive and the shape of the dative, the dative having replaced the genitive except in Greek, where the genitive replaced the dative. The same forms thus do double duty for marking possession and indirect objects. Romani and WRT maintain the genitive/dative distinction, and the situation is more complicated in Albanian and MR. Albanian has merged genitive and dative but has a distinct ablative. The dative is used as the object of a verb, the genitive is preceded by a particle of concord, and the ablative is the object of certain prepositions or in apposition to another substantive. In the indefinite plural, however, Albanian has a special ablative form in *-sh*. Pronominal declension also has a distinct ablative form used with certain prepositions, NOM *nga unë/djali* 'from me/the boy', ACC *për mua/djalin* 'for me/the boy', DAT *më tha mua/i tha djalit* 'he told me/the boy (with initial clitic reduplication)', ABL *prej meje/djalit* 'from me/the boy'. MR preserved a remnant of the genitive-dative distinction, albeit only in the speech of the oldest generation: *cari* 'who' *pe cari* 'whom.ACC', *la cari* 'to whom.DAT' but *al cruj* 'of whom, whose'. Elsewhere, the dative and accusative are distinct, and the genitive is identical to the dative.

**Analytic Case Relations** All the Balkan languages have simplified their inherited patterns of inflection. Eastern Macedonian and colloquial Bulgarian have gone the farthest, completely eliminating all traces of case morphology other than accusative personal pronouns and accusative vs. dative clitics. The marking of nonclitic dative objects is by means of the preposition *na* and the accusative pronoun. All other case relations are likewise indicated syntactically throughout BS, usually by a preposition but sometimes just by apposition. Western Macedonian preserves a distinctive set of dative synthetic pronouns, and, in the dialects that serve as the basis for the standard, a few remnants of animate singular masculine accusatives. As one moves further to the periphery of BS in the southwest and north, the complexity of case marking increases to include feminine accusatives, masculine datives, feminine datives, and eventually, in Gora and Torlakia, oblique plurals. In the Torlak dialects and the Macedonian dialects around Korça in Albania, case marking also occurs in the definite article. The other Balkan languages all retain at least three distinct cases (nominative, accusative, and genitive-dative).

Balkan Romani and WRT both preserve their full inflectional systems, but with tendencies toward simplification that show an intersection between the areal and typological. From a typological point of view, it is the peripheral cases that are expected to be lost first, and this is precisely what happens. Thus, WRT exhibits dative-locative confusion:

gitti-k	Selanik-te
went-1PL.AOR	Salonica-LOC
'We went to Salonica'	

There is also a tendency to eliminate case marking in locational postpositions:

ürti	üsti [vs üstü-n-de]	kedi-ler
blanket	top	cat-PL
'on top of his blanket	[there were]	cats'

Romani dialects in contact with BS tend to replace the locative with the dative and the dative, locative, and ablative with prepositional constructions derived from case affixes, themselves of postpositional origin:

jejh-e	aindž-a-te	vs. jekh-e	aindž-a-ke	> k-i jekh aindž
one-	field-	one-	field-	to-FEM one
OBL	OBL-	OBL	OBL-	field
LOC		DAT		
'in a field',		'to a field'		'in/to a field'
aindž-a-tar	= tar-i aindž			
field-OBL-ABL	from-FEM field			
'from a field'.				

Outside the pronouns, a distinct Romani accusative is limited to animate (or in some dialects referential) nouns, while in Turkish accusative marking is limited to definite or specific direct objects (see 'Resumptive Clitic Pronouns' [Reduplication, Replication]).

The vocative survives in all the Indo-European Balkan languages, and some argue that this preservation is a shared archaism, reinforced by contact, which is consistent with the direct encounters that lead to contact phenomena. It runs counter to the tendency toward analyticism, however.

**Analytic Gradation of Adjectives** Although the comparative is analytic in all the Balkan languages, remnants of synthetic comparatives survive at the peripheries, i.e., Greek has a number of inflected comparative forms, and northern Torlak preserves a very limited set. In the rest of BS, analytic comparatives with *po* are realized with almost complete consistency. Southern Montenegrin dialects also have analytic adjectival gradation using the same markers. BR, Albanian, and most Balkan dialects of Romani have complete consistency in the analytic marking of the comparative, the markers being *mai* (<*magis*) in Romanian and Megleno-Romanian, *cama* (*quam + magis*) in Aromanian, *më* in Albanian, and borrowed

in Romani (generally the marker of the main contact language, but Slavic *po* and Turkish *da/b/a* are both more widespread). Remnants of a synthetic comparative in *-eder* also survive in some Romani dialects, but generally those spoken outside the Balkans. Given that Romani entered the Balkans some time between the 10th and 13th centuries, and given that during this same period Slavic preserved its inflectional system of adjectival gradation, it would appear that BS and Romani were undergoing this shift at about the same time, and those dialects that left the Balkans did so before its completion.

In general, the standard of comparison is an ablative marker, which is synthetic in Turkish and most of Romani but prepositional (lexical 'from') in BS, BR, Greek, some Romani, and Albanian, particularly Tosk. Albanian can also use relative *se* and BR can have relative *ca*. Clausal comparisons (e.g., 'to eat is better than to sleep') in Albanian, BR, and BS involve quantifiers, *se [sa]* 'that [how much]', *de.cit* 'from. how much', *ot.kolko/fo* 'from.how much [that]', respectively. Greek has *pará* 'contrary to, despite'.

There is a bifurcation in the superlative between Turkish and BS, on the one hand, and Greek and Albanian on the other, with BR and Romani occupying a middle ground. In Turkish and BS, the relative superlative is purely analytic and uses native markers: Turkish *en*, BS *naj*. In Greek and Albanian, the relative superlative is expressed by the definite of the comparative. (Greek also has a synthetic absolute comparative in a few adjectives.) Romanian and most of MR pattern like Albanian, whereas Aromanian and the MR of Tsârnareka have borrowed Slavic *naj*.

The expression of analytic adjectival gradation in Turkish is attested in the oldest monuments (8th century). The Greek dialects of Epirus, Thrace, Asia Minor, and of the Sarakatsan (transhumant Hellenophone shepherds) use the comparative marker [*a/jkóm/a*] 'yet, still', calquing exactly Turkish *daha* (Table 1).

In Moldavian Gagauz, *sam* (<Russian *samyj*) is in competition with *en* as the superlative marker for the younger generation of speakers.

**Numerical Formation: The Teens** The formation of teens by means of a construction meaning 'numeral on ten' is pan-Slavic but absent from Baltic, occurs in BR but not the rest of Romance, and is also Albanian. Although assumed to be a calque from BS into BR and Albanian, Hamp (1992) has pointed out that the words for 'twenty' in BS and BR and 'thirty' in Albanian show the numeral 'ten' is masculine in Slavic but feminine in Albanian and BR. Based on the isomorphism in gender for BR and Albanian and a combination of old shared sound changes and ancient

**Table 1** Balkan adjectival gradation

Turkish	daha	büyük	ablative	en büyük
Romani (Arli)	po/da[h]a	baro	ablative	en/naj baro
Bulgarian	po-	goljam	ot	naj-goljam
Macedonian	po	golem	od	najgolem
Aromanian	kama	mari	di	nai mari
MR (Tsārnareka)	mai	mari	di	naimar[!]i most big
MR	mai	mari	di	tsäl mai mar[!]i
Romanian	mai	mare	de[cit]	cel mai mare
Albanian (Tosk)	më	i madh	nga	më i madhi
Greek	pio	megálos	apó	o pio megálos
	more	big	from	the more big 'biggest'
	'bigger than'			

**Table 2** Balkan teens and tens<sup>a-f</sup>

OCS	edinū	na	desete
Romanian	un	spre	zeci
Aromanian	ună	spră	[džatse]
MR	un	spră	ts
Albanian (Tosk)	një	mbë	dhjetë
	one	on	ten
Greek	enteka (en[a] deka)		
	one ten		
Romani	deš u jekh		
	ten and one		
Turkish	on bir		
	ten one		
	'eleven'		

<sup>a</sup>Slavic gender in numerals: dva (MASC) dve (FEM) 'two'.

<sup>b</sup>Romanian gender in numerals: doi (MASC) două (FEM) 'two'.

<sup>c</sup>Albanian gender in numerals: tre (MASC) tri (FEM) 'three' (dy [MASC], dý [FEM] 'two').

<sup>d</sup>OCS 10 = MASC düvä desete 'twenty'.

<sup>e</sup>Romanian 10 = FEM două zeci 'twenty' (zece 'ten' < Lat. decem).

<sup>f</sup>Albanian 10 = FEM tri dhjetë 'thirty'.

borrowed lexicon among the three, Hamp suggests that this innovation occurred at a time when the Indo-European dialects that became Slavic, Albanian, and the language that Latinized into Romanian were part of a Northwest European sprachbund prior to their respective migrations to the Balkans (Table 2).

**Analytic Subjunctive** The analytic subjunctive formed by means of a subordinating particle (SP), usually of pronominal origin, plus a finite verb agreeing with its subject (omitted if the same as in the main clause, specified if different) replaces older nonfinite complements (infinitives) in all Balkan languages to varying degrees. Gheg has a new infinitive employing the preposition *me* 'with' and a short participle in contexts where Tosk uses the analytic subjunctive, but Gheg also has uses of the analytic subjunctive, and Tosk has some nonfinite participial constructions where

other Balkan languages have the analytic subjunctive. Romanian and MR still have remnants of the Latin infinitive that can be used in some traditional infinitival functions. The BR infinitive is strongest in Maramureş, the northernmost Romanian region and the one in most contact with infinitive-using languages (Ukrainian, Hungarian, formerly Yiddish). BR in general also preserves Latin infinitives in *-re* as verbal nouns. Greek has a morphological remnant of the infinitive, but its only living function is to represent the main verb in perfects and pluperfects. Bulgarian has a very marginal remnant of the Slavic infinitive limited to subordination to a tiny number of verbs. The infinitive has disappeared completely from Torlak except in some folk songs. Macedonian and Romani have eliminated all traces of earlier infinitives. Thus the replacement of infinitives with subjunctives is not uniform but scalar. At one end is Gheg, followed closely by Romanian, then Tosk, Bulgarian, Greek, and Vlah, with Torlak, Romani, and Macedonian at the other end.

New infinitival constructions have arisen in Romani outside of the Balkans in contact with infinitive-using languages. In Macedonian, some uses of the verbal noun can replace SP-clauses and thus function as a kind of new infinitive, although these constructions, which are highly colloquial, are merely alternatives. The option of using an SP-clause rather than an infinitive is available to all of BCS, but there is a tendency for such usage to become more frequent as one moves from northwest to southeast in the direction of Torlak. Since 1991, Croatian language planners have identified SP-clauses with Serbian and infinitives with Croatian, as a result of which Croatian speakers are now discouraged from using SP-clauses. In Serbian and Bosnian, however, the two constructions continue to coexist amicably (Table 3).

In WRT, optatives have expanded at the expense of infinitives owing to the influence of the other

**Table 3** Balkan SP clauses

Romani	mangav	te	hramonav
Albanian (Tosk)	dua	të	shkruaj
Albanian (Gheg)	[due	me	shkrue]
Greek	thélo	na	gráfō
Bulgarian	iskam	da	piša
Macedonian	sakam	da	pišuvam
Torlak	oču	da	pišem
Romanian	vreau	să	scriu
Vlah	voi	s(i)	scriu
<i>gloss</i>	<i>I.want</i>	<i>SP</i>	<i>I.write</i>
WRT	isterim		yazayım
<i>gloss</i>	<i>I.want</i>		<i>I.write.OPT</i>
			'I want to write'

Balkan languages. The usage in **Table 3** was a possibility in older Turkish, but, in a classic case of convergence via feature selection, the WRT optative now occurs where Turkish would normally have a nonfinite construction:

ben seni ist-er-im şimdi bir  
*I you.ACC want-PRES-1SG now one*  
 müneccim ol-a-sin  
*astrologer be-OPT-2SG*  
 Now I want you to be an astrologer

Similarly, Balkan Judezmo, which preserves the Spanish infinitive, nonetheless has some uses of its subjunctive, e.g., in questions, that calque Balkan SP-clauses and do not occur in Modern Spanish or North African Judezmo:

kwando ke te vengamoz a tom-ar?  
 (Balkan Judezmo)  
*when that you.ACC we.come to take-INF*  
 Póte na 'rthoúme na se pároume? (Greek)  
*when SP we.come SP you.ACC we.take*  
 Koga da ti dojdeme da te zemame?  
 (Macedonian)  
*when SP you. we.come SP you. we.take*  
 DAT ACC  
 Cuándo quieres que vengamos a recog-er-te?  
 (Modern Spanish)  
*when you.want that we.come to take-INF-you*  
 'When do you want us to come to get you?'

All Balkan languages use the independent analytic subjunctive to express wish, desire, or a milder form of imperative. Albanian also has a synthetic optative used mostly in formulae.

**Futures in 'Will' and 'Have'** When Slavic entered the Balkans (6th–7th centuries C.E.), there was competition between the auxiliaries 'have' and 'want (will)' + infinitive to mark futurity in Latin and Greek, with Latin favoring 'have' and Greek favoring 'want'. OCS used the perfective of 'be' in addition to 'want', 'have' and various forms of 'begin' + infinitive.

**Table 4** Balkan futures

Romani	ka	dža[s]
Albanian (Tosk)	do	[të] shkojmë
Greek	tha	páme
Bulgarian	še	trügnem
Macedonian	ke	odime
Torlak	če	odime
Romanian (Colloquial, South)	o	să mergem
Aromanian	va	s- neadzim
MR		si, să neadzim
[MR-Tsārnareka	ās	neadzim]
English		<i>we will go</i>

The 'will' + infinitive construction survives (with modified or new infinitives) in Romanian, northwest Gheg (near and in Montenegro), in Bulgarian dialects (with postposed auxiliary), and MR (for speculations and threats). This form also survives in all the non-Balkan Štokavian dialects of BCS and connects them with East South Slavic. In fact, much of Štokavian ended up in its current location as a result of northward migrations during the 15th–18th centuries. The rest of Slavic developed the perfective of 'be' as a future marker. The next stage was 'will' + SP + conjugated present tense verb (Greek 14th century, Slavic 15th century). This stage also survives in BCS, including Torlak. The third stage, which overlaps the second, is the transformation of 'will' into an invariant particle + SP + conjugated main verb. This type of construction is still the main one in Tosk and parts of Gheg, especially in the northwest and southeast peripheries; it is characteristic of southern Romanian and survives in Torlak and in certain modal uses in East South Slavic, but not in Greek. The fourth stage is the elimination of the SP so that the future is marked by an invariant particle plus a conjugated verb. In addition to being the standard future in Balkan and southern Vlax Romani, Greek, and BS, it is common in colloquial Tosk. In MR, the future marker merged with SP, producing a new particle, ās, in Tsārnareka, but eliminating a distinct future marker in the other villages. Romani outside the Balkans has other means of forming or expressing the future, and it appears that the Romani development in the Balkans occurred in concert with the other Balkan languages (cf. 'Analytic Gradation of Adjectives') (**Table 4**).

Conjugated 'have' + infinitive, attested for early stages of all the traditional Balkan languages, remains the predominant future in most of Gheg. Conjugated have + SP + present is still used in Romanian, and invariant 'have' (which can also be an existential in all the Balkan languages with lexical 'have', cf. French *il y a*) is used in Arbëresh and occurs with modal

**Table 5** Negated futures

Macedonian	nema	da	odime
Bulgarian	njama	da	hodime
Aromanian	noare	s'	neadzim
	not.has	SP	we.go
Romani	na-e	amen	te dža[s]
	not-is	we.ACC	SP go.IPL.PRES
WRT	yok-tur		gid-elim
	not-is		go-OPT.1PL
English	we won't go		

functions in BS. In East South Slavic, the ordinary negated future uses this negative existential + SP + present, and this type is calqued into Aromanian, Romani, and WRT. Since Turkish and most of Romani lack lexical verbs meaning ‘have’, their calques use their negated existentials, which also code possession (**Table 5**).

**Future in Past as Conditional** The combination of a future marker with a past tense marker to form a conditional, especially irrealis, is a classic Balkanism, although its realization differs among the various Balkan languages. (The construction itself can have a variety of related meanings, e.g., ‘X almost happened/was about to happen’, iterative-habitual, anterior future, and languages and dialects can be differentiated on the basis of which of these meanings are encoded.) Greek, Macedonian, and Romani all use the invariant future marker plus the imperfect. Tosk and Aromanian are almost the same, but they still have the SP, at least optionally. MR has an invariant ‘will’ marker (*vreă*) + SP + present or perfect (see ‘Perfect in “Have”’). In Bulgarian, Torlak, and other dialectal BCS, however, it is ‘will’ that conjugates in the imperfect+SP+present, and Gheg has the conjugated imperfect auxiliary ‘have’ + infinitive. The Balkan construction extends into BCS as far as southern Croatia and southwestern Serbia, and the southern Montenegrin dialects have the widest range of uses for the construction, thereby being most Balkan. In Turkish, the future participle plus a past auxiliary [*i*]di or [*i*]miş has the same nuances of irrealis conditional (**Table 6**).

In Greek, Albanian, and Vlah, conditional constructions normally have a form of the ‘will’ morpheme. In BS, the Balkan conditional is in competition with the inherited conditional using the old optative of ‘be’ (invariant *bi* in Macedonian, conjugating in Bulgarian and Torlak) + old resultative participle. Romani dialects in contact with Slavic also use invariant *bi* + present as a conditional. In Romanian, a special conjugation of ‘have’ + infinitive serves as a conditional-optative.

**Table 6** The Balkan conditionals

Romani	ka	keravas*
Greek	thá	égrafa
Macedonian	ke	napravev <sup>u</sup>
Aromanian	va	făceam <sup>u</sup>
Albanian (Tosk)	do	të bëja
	FU	SP do.IM.ISG
MR	vreă	si am fat(ā)
	want.PRES.3SG	SU do.PERF.1SG
Bulgarian	štjah	da napravja
	ščaše/ščeše	da napravim/radim
	want.3SG.IM	SU do.PRES.1SG
Albanian (Gheg)	[kishna I.have	me bâj do.PART
Romanian	aş COND	fi făcut
Turkish	yap ROOT	be.INF acak tı m PAST 1SG
		‘I would have done’

\*Arli has a new imperfect formed by the long present + imperfect of ‘3SG/PLbe’, e.g. *kerava sine*.

**Perfect in ‘Have’** The use of ‘have’ as an auxiliary with a nonfinite main verb to form an analytic perfect is attested for Greek and Latin at the end of the ancient period and is characteristic of Albanian, BR, and Greek, while such constructions (and lexical ‘have’) are absent from WRT and most of Romani. In BS ‘have’ + past passive participle (or its descendant) forms resultative constructions ranging from a fully grammaticalized perfect (with an invariant neuter verbal adjective) that has completely replaced the inherited perfect (‘be’ + old resultative particle in *-l*) in extreme southwestern Macedonian and spreading north to Mt. Šar and east to the Vardar and beyond, to resultative syntagms with ‘have’ + past passive participles agreeing with their direct objects and limited to transitive verbs with human subjects in most of Bulgaria.

Given the geography and history of the ‘have’ perfect in BS, it is clearly a calque on one of the non-Slavic contact languages. Although Greek and Albanian have been proposed as the possible models, Gołęb's arguments in favor of Aromanian are the most convincing. In Aromanian, the feminine participle is selected as the invariant, since in BR (as in Albanian) the feminine gender is unmarked (neuter is obsolete). The Macedonian invariant neuter verbal adjective therefore corresponds exactly to the Aromanian in terms of unmarked gender. In Greek, the main verb is a remnant of the infinitive and in Albanian the participle does not mark singular gender. Thus the BR construction most closely resembles the Macedonian. An additional argument in favor of BR as the model is evidence of Macedonian and Vlah mutual calquing in other resultative constructions.

**Table 7** Aromanian (Fărshălots, Bela di Supră and Albanian indicatives (3sg 'work')

	<i>Nonadmirative</i>		<i>Admirative</i>	
<i>Present</i>	lukră	punon	lukracka	punuaka
<i>Perfect</i>	ari lukrată	ka punuar	avuska lukrată	paska punuar
<i>Pluperfect</i>	ave lukrată	kish punuar	—	paskësh punuar
<i>2nd Pluperfect</i>	avu lukrată	pat punuar	—	—
<i>Double perfect</i>	ari avut lukrată	ka pasē punuar	ari avuska lukrată	paska pasē punuar
<i>Double plup.</i>	ave avut lukrată	kish pasē punuar	ave avuska lukrată	paskësh pasē punuar
<i>2nd Dbl. plup.</i>	avu avut lukrată	pat pasē punuar	—	—

**Table 8** The Novo Selo probabilitive 'see'

Present	1	gla*dàčă m	gla*dàčă mo
	2	gla*dàčăš	gla*dàčă tă
	3	gla*dàčă	gla*dàčă ju
Future		čă gla*dàčă m, etc.	
Past		budàčă m ~ bìčă m glă dàl, etc.	

**Evidential** In a Balkan context, evidentiality (inferential, distance, mode of indirect narration, indirective, status, French *médiatif*) is a grammatical category encoding the speaker's evaluation of the narrated event, often, but not always, predicated upon the nature of the available evidence. Evidentials can be of two types: confirmative (vouched for, 'witnessed') and nonconfirmative (not vouched for, 'reported', 'inferential'). The nonconfirmative can be felicitous (neutral report or inference) or infelicitous, in which latter case the nonconfirmative expresses either acceptance of a previously unexpected state of affairs (i.e., surprise, admirativity *sensu stricto*) or rejection of a previous statement (i.e., sarcasm, dubitativity). The opposition confirmative/nonconfirmative was already encoded in the Turkic simple past in *-di* (confirmative) and the perfect participle in *-miş* (nonconfirmative) at the time of the earliest monuments. In East South Slavic, the old synthetic pasts are markedly confirmative (this same meaning is also sometimes identified in Torlak). By contrast, the old perfect using the resultative participle in *-l* has become an unmarked past, with a chief contextual variant meaning of nonconfirmative. In Albanian, the inverted perfect (participle + 'have') has fused into a marked nonconfirmative present paradigm called admirative, which can then function as an auxiliary to form analytic past tenses. The Frasherote Aromanian dialect of Bela di Supră has reinterpreted the 3SG.PRES Albanian admirative marker as an admirative suffix, which it adds to a masculine plural imperfect participial base to form a new admirative (Table 7).

Megleno-Romanian uses an inverted perfect + auxiliary construction in a similar function. The Romanian presumptive mood formed with a future, subjunctive, or conditional marker + invariant *fi* 'be' + gerund (or past participle) is a similar marked nonconfirmative, as is the probabilitive mood (based on a BCS-type inverted future) of Novo Selo Bulgarian, a dialect spoken across the Danube from Romania and a few kilometers east of Serbia (Table 8).

The Judezmo of Istanbul uses the pluperfect as a calque on the Turkish past in *-miş*:

Kuando <i>when</i>	esta-v-an <i>be-IM-3PL</i>	en l' Amérika, les <i>in the America them.DAT</i>
av-i-y-a <i>have-IM-3SG</i>	entra-do <i>enter-PAST.PART</i>	ladrón <i>thief</i>
'When they were in America [i.e., absent], a thief broke into (Turkish <i>girmış</i> ) their house.'		

**Other** Many other features too numerous to discuss here are cited as Balkanisms, e.g., the conflation of adverbs of location and motion ('where'/'whither'), purposives in 'for' + SP + verb and other prepositional parallelisms, a distinction between realis and irrealis complementizers, and absolute relativizers and interrogatives as complementizers, this last being a feature that has spread to WRT:

čovek-ot	što	go	vid-ov	(Macedonian)
<i>person-the</i>	<i>what</i>	<i>him.ACC</i>	<i>see-1SG.AOR</i>	
adam	ne	cür-d-üm	(WRT)	
<i>man</i>	<i>what</i>	<i>saw-PAST-I</i>		
gör-düg-üm	adam	(Standard Turkish)		
<i>see-PART-my</i>	<i>man</i>			
'the man that I saw'				

### Word Order

**Ctic Ordering** Greek, Albanian, and BR all permit absolute initial pronominal clitics when the first stressed element is a finite verb, but in BS only Macedonian (especially the western dialects) permits this. Bulgarian keeps pronominal clitics bound to the verb

but either requires the verb or some other element in initial position. BCS, including most of Torlak, still follows Wackernagel's law and has clitics in second position.

Az često mu	go	davam. (Bulgarian)
I often. him.DAT	it.ACC	give.1SG.PRES
Ja mu	ga	često dajem. (Serbian)
I him.DAT	it.ACC	often give.1SG.PRES.
'I often give it to him'		
Davam	mu	go (Bulgarian)
give.1SG.PRES	him.DAT	it.ACC
Mu	go	davam (Macedonian)
him.DAT	it.ACC	give.1SG.PRES
'I give it to him'		

Romani pronominal clitics follow the verb. WRT is basically suffixal, like the rest of Turkish, and clitics always follow the stressed element, but elements that can be fused or separated are more likely to be separated and less likely to show vowel harmony in WRT.

**Constituent Order** Balkan languages are characterized by relatively free constituent order with certain patterns being favored for various types of syntactic and narrative strategies (emphasis, topicalization, focus, contrastive thematization, etc.). The unmarked word order tendency is SVO in all the Indo-European Balkan languages. Unlike most of Turkish, where the tendency is verb-final, WRT and Gagauz show SVO tendencies. Similarly, BR, BS, Albanian, and Greek all have the basic order head-genitive, while Turkish and Romani are genitive-head. Romani dialects in the Balkans and WRT, however, also have head-genitive constructions:

m-e	phral-es-k(er)e	kher-es-k(or)o	vudar	(Romani)
my-OBL	brother-OBL-	house-OBL-	door	
GEN	GEN			
'the door of my brother's house'				
o	vudar	e	kher-es-ko	
the.MASC.NOM	door	the.OBL	house-OBL-GEN	
m-e	phral-es-kere	(Romani)		
my-OBL	brother-OBL-GEN			
Baba-si	Ali-nin	(WRT)		
father-his	A.-GEN			
Tatko	mu	na	Ali	(Macedonian)
father	him.DAT	to	Ali	
Baba-i	i		Ali-ut	(Albanian)
father-DEF	PC.MASC.NOM.SG		A.-Def.GEN	
Ali-nin	babasi	(Standard Turkish)		
Ali-GEN	father-his			
'Ali's father'				

Adjectives generally follow their heads in Albanian and BR, but precede in BS, Greek, WRT, and Romani. In all of these languages, the opposite order is possible in various discourse functions. Albanian enclaves in

the eastern Balkans also have preposed adjective as the standard order.

### Lexicon, Semantics, and Derivational Morphology

The etymological commonalities of the Balkan lexicon received considerable attention during the formative years of Balkan linguistics, whereas more recently the focus has been on shared grammatical features. Miklosich's 1861 survey of Balkan grammatical commonalities occupied only 4% of what was basically a study of the Slavic lexical influence on Romanian. Sandfeld (1930) devotes 40% of his book to the lexicon, whereas Asenova (2002) allots 10% of her text to such issues. Although the lexicon is the most salient surface manifestation of linguistic influence, words can travel between languages without the aid of communal multilingualism, whereas the diffusion or convergence of grammatical structures is a more complex process that requires at least a core community of bi- or multilingual speakers. In terms of the definition of a sprachbund, it is the shared grammatical features rather than shared vocabulary that is the key determiner, although shared vocabulary is usually part of the picture.

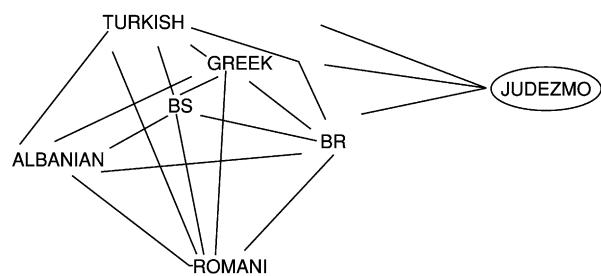
There are common loanwords from each of the component language families in the Balkan languages. Words shared by Albanian and Romanian of pre-Latin (substrate) origin are often connected with domestic items or husbandry, e.g., Albanian *shtrungë*, BR *strungă*, BS (Macedonian and west Bulgarian) *strunga*, Greek (Epirus and Sarakatsan) *stroúgka* 'dairy'. Greek, Slavic, and Romance (especially Balkan Latin and Venetian Italian) were all languages of power in the Balkans at various times during the Middle Ages and contributed a variety of lexemes and even derivational affixes to the common Balkan lexicon, e.g., the Latin agentive suffix *-arius*, the Slavic feminine suffix *-ica*, and the Greek aorist marker *-s-* (used in deriving verbs). As the language of administration, the market place, and urban life in general, Turkish dominated the Balkan peninsula for more than half a millennium. By the 19th century, the shared Turkish lexicon in the Balkan languages was of considerable size. The rise of Balkan standard languages, however, entailed the stylistic lowering and marginalization of many Turkish loanwords, and as many of these items were of Arabo-Persian origin, they were discouraged by Turkish purists as well. The Turkish agentive *-ci*, attributive *-li*, qualitative or concrete *-lik* (with adjustments for vowel harmony, voicing assimilation, and adaptation) continue to be productive as derivational affixes, e.g., Macedonian *puberte-tlija* 'adolescent (ironic)' Albanian *partiakçi* 'party hack', Judezmo *hanukalik* 'Chanukah present', etc.

The Balkan languages also share numerous idioms, collocations, and calqued expressions; e.g., the use of ‘eat’ to mean ‘undergo something unpleasant’ as in ‘eat wood’ = ‘take a beating’ or ‘it doesn’t cut his mind’ = ‘he doesn’t understand’. There are a variety of shared discourse particles and conjunctions (e.g., Turkish *am/mja* ‘but’, Greek *bre* ‘hey, vocative particle’) that also form part of the common Balkan lexicon.

## Sociolinguistics

Factors such as power, prestige and religion have influenced directions and degrees of Balkan contact phenomena. Throughout the Ottoman period, Turkish had high prestige as the language of the state and the town, Greek had prestige among Christians as the language of the (Orthodox) church with its own literary tradition (and history of power, i.e., Byzantium) and was also a language of commerce. BS had less prestige in the southern Balkans, but its history of medieval literacy and political competition with Byzantium gave it some limited prestige. Although BR was descended from Latin, another language of empire and conquest, the local varieties that developed after the Slavic invasions did not have that level of prestige and, like Albanian, were associated mainly with rural contexts. In Wallachia and Moldavia, Church Slavonic was the liturgical language for centuries, and Romanian was written in Cyrillic until the mid-19th century. Aromanian speakers in southern Balkan towns used Greek outside the home. Romani was at the bottom of the social hierarchy, but Judezmo was outside it. This is reflected in 19th-century Macedonian folklore collections, where characters in ethnic jokes, including Roms (Gypsies), speak in their own languages, except Jews, who speak Turkish, not Judezmo. For both Romani and Judezmo, multilingualism was unidirectional, i.e., Roms and Jews learned other languages but heard their languages spoken by others rarely, if ever. At the opposite end of the prestige scale, speakers of Greek and Turkish were less likely to learn less prestigious languages but were more likely to hear their languages spoken by others. Those languages in the middle of the hierarchy (BS, BR, and Albanian) had the highest degree of multidirectional multilingualism and show a higher degree of congruence.

Marriages could be freely contracted across linguistic lines but not religious ones, so that multilingual households were a commonplace. Although speakers of BS, BR, and Greek were mostly Christian and speakers of Albanian were usually Muslim, each of these religions also had significant communities speaking the other languages. Except for Gagauz,



**Figure 1** Schematic linguistic social/political hierarchy (Ottoman Period).

speakers of Turkish were Muslim, but there was still plenty of linguistic contact via religious conversion. Jews and Roms, however, were endogamous along a combination of linguistic and other social lines. This boundary maintenance is reflected linguistically in Romani, where there is a clear opposition between the relatively open systems of adjectival comparison and modality on the one hand to the conservative nominal, pronominal, and tense-aspect systems on the other.

**Figure 1** illustrates the relative prestige of the various languages during the Ottoman period. Height symbolizes prestige, while incline indicates relative (never absolute) directionality. The directionality is reversed in the case of slang and secret languages, where it is the covert prestige of languages further down on the social scale that is reflected in patterns of lexical borrowing. In the case of Judezmo, knowledge of Turkish was most widespread, while knowledge of other Balkan languages would depend on the particular (urban) environment.

## Causation

For most of the history of Balkan linguistics, causation has been sought in the influence of (interference from) one of the languages, e.g., Greek, Latin, or a pre-Latin non-Hellenic substratum (e.g., Illyrian, Thracian, and/or Dacian – all so poorly attested that we do not have so much as a single sentence in any of them). More recently, however, an ecological model of feature selection argues that those grammatical developments more suitable for effective communication that might be already present in the language, i.e., more adaptive, are more likely to be selected for further development and spread (cf. ‘Resumptive Clitic Pronouns [Reduplication, Replication]’). In such a model, languages can utilize native resources that are reinforced by their occurrence, or potential for occurrence, in the contact languages. Mechanisms such as fusion, metatypy, and code copying are all potentially relevant. At the same time, sociolinguistic factors such as those adduced in ‘Sociolinguistics’ can

influence directions of change. The diffusion of borrowings and the development of convergences are thus compatible parts of a larger picture of a sprachbund in which languages come to be similar without becoming identical. It is worth emphasizing here the insight of Joseph (2001), namely that the move from lexical via phraseological to syntactic borrowings that characterizes the contact-induced changes of a sprachbund such as the Balkans are quintessentially surface phenomena.

Although some scholars have argued against the idea of a Balkan sprachbund since the 1930s, the argument that the Balkans are basically just part of a larger European linguistic zone coincides roughly with the recent rise in interest in contact linguistics and typology. In the case of the Balkans, however, while it is clear that Kopitar's formulation is an exaggeration, it is equally clear that Trubetzkoy's original insight captures facts about language relationships. Of particular significance is the manner in which patterns map such that the languages that surround the Balkan sprachbund do not share the most salient features. The fact that English and Western Romance have gone even further than most of the Balkan languages in some changes does not contradict the hypothesis that the Balkan sprachbund is precisely that, i.e., a product of the process of language contact. If some of those contact-induced changes are the result of shared feature selections, having parallels elsewhere, that may contribute to identifying likely directions of language change, but it does not vitiate the sprachbund as a historical and sociolinguistic phenomenon.

In a sense, a sprachbund is more like a dialect chain than a linguistic family: as features spread over areas, they may do so with differential impact. Thus, while it is possible to define a sprachbund in terms of languages displaying a coalescence of a number of such features, it is not necessarily the case of an 'all and only' phenomenon. Moreover, the transition from pragmatic to syntactic (grammaticalized) to morphological sometimes maps onto the territory of the sprachbund itself, moving from periphery to core. Like dialects, there can be a transitional effect, and a given language, e.g., BCS, can participate in the changes to a greater or a lesser extent. For both the dialect and the sprachbund, politics can have a crucial effect in setting boundaries that favor internal consistency and external differentiation. Just as the very concept of language vis-à-vis dialect (e.g., to which language a given dialect 'belongs' or which isoglosses will be chosen as defining one dialect in opposition to another) can be a complex of intersecting factors, so too can the definition of sprachbund.

*See also:* Africa as a Linguistic Area; Albanian; Albania: Language Situation; Areal Linguistics; Articles, Definite and Indefinite; Bilingualism; Bosnia and Herzegovina: Language Situation; Bulgaria: Language Situation; Bulgarian; Case; Clitics; Contact-Induced Convergence: Typology and Areality; Ethiopia as a Linguistic Area; Europe as a Linguistic Area; Evidentiality in Grammar; Greece: Language Situation; Greek, Ancient; Greek, Modern; Language Change and Language Contact; Language Politics; Latin; Macedonia: Language Situation; Macedonian; Metatypy; Mood and Modality in Grammar; Nationalism and Linguistics; Old Church Slavonic; Romani; Romanian; Romania: Language Situation; Sanskrit; Serbian–Croatian–Bosnian Linguistic Complex; South Asia as a Linguistic Area; Southeast Asia as a Linguistic Area; Turkey: Language Situation; Turkish.

## Bibliography

- Alexander R (2000). 'Tracking sprachbund boundaries: word order in the Balkans.' In Gilbers D, Nerbonne J & Schaeken J (eds.) *Studies in Slavic and General Linguistics 28: Languages in contact*. Amsterdam: Rodopi. 9–27.
- Asenova P (2002). *Balkansko ezikoznanie. Osnovni problemi na balkanskija ezikov sajuz* (Balkan linguistics: Basic problems of the Balkan linguistic league) (2nd edn.). Sofia: Farber. (Resumé in English, French, and Russian).
- Atanasov P (1990). *Le mélénoroumain de nos jours*. (Balkan-Archiv Neue Folge – Beiheft 8). Hamburg: Buske.
- Belyavski-Frank M (2003). *The Balkan conditional in South Slavic*. Munich: Otto Sagner.
- Friedman V (2003a). 'Evidentiality in the Balkans.' In Aikhenvald A & Dixon R (eds.) *Studies in evidentiality*. Amsterdam: Benjamins. 189–218.
- Friedman V (2003b). *Turkish in Macedonia and beyond: studies in contact, typology, and other phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz.
- Friedman V (2004). *Studies on Albanian and other Balkan languages*. Peja: Dukagjini.
- Friedman V & Joseph B (2006). *The Balkan languages*. Cambridge: Cambridge University.
- Golab Z (1984). *The Arumanian dialect of Kruševo in SR Macedonia SFR Yugoslavia*. Skopje: Macedonian Academy of Arts and Sciences.
- Golab Z (1997). 'The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian.' *Balkanistica* 10, 13–19.
- Gălăbov I, Georgiev V & Zaimov J (eds.) (1968). *Actes du premier congrès international des études balkaniques et sud-est européens* (vol. 6). Sofia: Bălgarska Akademija na Naukite.
- Hamp E (1982). 'The oldest Albanian syntagma.' *Balkansko Ezikoznanie* 25, 77–79.
- Hamp E (1992). 'Albanian.' In Gvozdanović J (ed.) *Indo-European numerals*. Berlin: Mouton de Gruyter. 835–921.

- Hinrichs U (ed.) (1999). *Handbuch der Südosteuropa-Linguistik*. Wiesbaden: Harrassowitz.
- Ilievski P (1988). *Balkanološki lingvistički studii* (Balkanological linguistic studies). Skopje: Institut za makedonski jazik.
- Joseph B (1983). *The synchrony and diachrony of the Balkan infinitive: a study in areal, general, and historical linguistics*. Cambridge: Cambridge University Press.
- Joseph B (1987). 'A fresh look at the Balkan Sprachbund: some observations on H. W. Schaller's *Die Balkansprachen*.' *Mediterranean Language Review* 3, 105–114.
- Joseph B (2001). 'Is Balkan comparative syntax possible?' In Rivero M & Ralli A (eds.) *Comparative syntax of Balkan languages*. Oxford: Oxford University Press. 17–43.
- Katičić R (1976). *Trends in linguistics, state-of-the-art report 4, 5: ancient languages of the Balkans* (2 vols). The Hague: Mouton.
- Koneski B (1967). *Gramatika na makedonskot literaturnen jazik* (Grammar of the Macedonian literary language). Skopje: Kultura.
- Lindstedt J (2000). 'Linguistic Balkanization: contact-induced change by mutual reinforcement.' In Gilbers D, Nerbonne J & Schaeken J (eds.) *Studies in Slavic and General Linguistics 28: Languages in contact*. Amsterdam: Rodopi. 231–246.
- Matras Y (1994). 'Structural Balkanisms in Romani.' In Reiter N (ed.) *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*. Wiesbaden: Harrassowitz. 195–210.
- Nikolaeva T (1996). *Prosodija Balkan: Slovo – vyskazyvanija – tekst* (Prosody of the Balkans: Word – Utterance – Text). Moscow, Nauka.
- Reiter N (ed.) (1983). *Ziele und Wege der Balkanlinguistik* (Balkanologische Veröffentlichungen, Vol. 8). Berlin: Osteuropa Institut an der Freien Universität Berlin.
- Sandfeld K (1930). *Linguistique balkanique*. Paris: Klincksieck.
- (1926). *Balkanfilologien*. Copenhagen: Luno.
- Sawicka I (1997). *The Balkan Sprachbund in light of phonetic features*. Toruń: Energeia.
- Schaller H (1977). *Bibliographie zur Balkanphilologie*. Heidelberg: Carl Winter.
- Sobolev A (2005). *Malyj dialektologičeskij atlas balkanskih jazykov*. (Small dialectological atlas of the Balkan languages). Marburg: Biblion Verlag.
- Solta G (1980). *Einführung in die Balkanlinguistik mit besonderer Berücksichtigung des Substrats und des Balkanlateinischen*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Topolińska Z (1995). 'Convergent evolution, creolization and referentiality.' In Hajičová E et al. (eds.) *Prague Linguistic Circle Papers*, vol. I. Amsterdam: Benjamins. 239–247.
- Varol M-C (2001). 'Calques morphosyntaxiques du turc en judeo-espagnol: mécanismes et limites.' In Donabedian A (ed.) *Faits de langues: langues de diaspora*. Paris: Ophrys. 85–99.

## Balochi

**P O Skjærø**, Harvard University, Cambridge, MA, USA

© 2006 Elsevier Ltd. All rights reserved.

Balochi (or Baluchi, in several dialects) is spoken by the Baloch in eastern Iran and western Pakistan (Baluchistan), but also in southern Afghanistan, Turkmenistan, and the Arab Gulf States (totaling 6–8 million speakers?). The Baloch are first mentioned in literature about 1000 C.E., but the language did not become a written one until the 20th century, although the earliest known manuscript dates from the early 19th century. On the other hand, the Baloch have an oral poetic tradition with historical themes reaching back to the 15th century, but especially productive in the 19th century. Modern literature and publications are centered in Quetta in Pakistani Baluchistan and in Karachi. A Balochi Academy was founded in Quetta in 1959 and still publishes Balochi literature and supports Balochi language and culture in various ways, and the University of Quetta offers a Balochi Studies program. Balochi radio programs are

broadcast from Zahedan in Iranian Balochistan and from Quetta and Karachi, formerly also from Kabul.

There are, by one count, six principal dialects of Balochi, characterized by differences in grammar and lexicon. The western dialect of Raxšānī is the largest, the principal subdialect being Sarhaddī,

Balochi belongs with the North(west) Iranian languages, differently from Persian, which is a Southwest Iranian language; compare, for instance, Balochi *asín* 'iron,' *jan-* [dʒan-] 'strike,' *zird* 'heart,' versus Persian *āhan*, *zan-*, *dil*. It is a phonetically conservative language, having preserved much of the Old Iranian consonant system intact, notably intervocalic stops and affricates, for instance, Bal. *pād* 'foot,' *āp* 'water,' *roč* [rōtf] 'day' (Pers. *pā*, *āb*, *rūz*). Among innovations are the development of initial *w-* to *g(w)* (OIran. *wāta-* 'wind,' Bal. *gwāt*, Pers. *bād*), *xw-* to *w-* (OIran. *xwara-* 'eat,' Bal. *war-*, Pers. *x<sup>v</sup>or-*), and the change of fricatives into stops (Bal. *nākun* 'nail,' Pers. *nāxon*; Bal. *gipta* 'seized,' Pers. *gereft*).

Balochi has retroflex consonants in words borrowed from Indo-Aryan, including originally English words, for instance, *drēwar* (*d* = [ɖ]) 'driver.'