# 4 Lexicon and Semantics

## 4.0 Introduction

This chapter treats the most particularly evident outcome in language contact, namely developments involving various sorts of lexical material. Included in this are a consideration of lexical semantics in the Balkans under conditions of contact and the identification of a particular type of loanword of great importance in the sprachbund.

## 4.1 On the Nature of Balkan Lexical Evidence and Lexical Evidence in General

Shared vocabulary is the most obvious manifestation of language contact, and even with a considerable amount of attention in the Balkanological literature to morphosyntax, the importance of the lexicon for Balkan studies, especially in the earlier days of the field, is clear. Lexical parallels were among the Balkan features noted by Miklosich 1862 (see §2.2.3) and, as observed in §3.4.2.1, so-called culture words were a basic part of Trubetzkoy's original conceptualization of a sprachbund.[1] Moreover, Sandfeld 1930 devotes nearly half his work to loanwords as well as parallels in phraseology, something he considers to be "*en dehors du lexique*" ('outside of the lexicon'), a characterization – consistent with views of the lexicon at the time he was writing – that depends on a narrow definition of "lexicon" as just involving words (lexical items). Moreover, one can look for confirmation of the value of the lexicon by examining the distributions of coverage in accounts of sprachbund convergences seen in various relatively recent handbooks of Balkan linguistics; the number of pages devoted to the lexicon as opposed to other domains offers an interesting perspective on the importance accorded the lexicon, as seen in Table 4.1.

These numbers indicate the importance given to coverage of the lexicon in Balkan language contact, but, at the same time, they support Kahl's 2014 conclusion that the lexicon has been relegated to reduced importance vis-à-vis morphosyntax in recent studies. We, too, consider the lexicon to be important, though as the material presented here shows, for somewhat different reasons.[2] Our focus here is on lexical aspects of language contact in the Balkans, not on the lexicon in the individual languages as separate synchronic systems. In keeping with this focus, our treatment of semantics is linked to word- and phrase-meaning and not to other

1 See §4.2 for an elaboration of this view.
2 See especially §4.3 on conversationally based "ERIC" loans.



Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.1 *Topic distribution in Balkan handbooks 1975–2012*

| WORK | PHONOLOGY (# PAGES) | MORPHOSYNTAX | LEXICON | %LEXICON | % PHONOLOGY |
|---|---|---|---|---|---|
| Asenova 2002 | 15 | 216 | 33 | 13% | 6% |
| Banfi 1985 | 5 | 31 | 31 | 46% | 7% |
| Sh. Demiraj 2004 | 12 | 76 | 12 | 12% | 12% |
| Feuillet 1986 | 9 | 37 | 13 | 22% | 15% |
| Feuillet 2012 | 22 | 156 | 29 | 14% | 11% |
| Schaller 1975 | 10 | 38 | 19 | 28% | 15% |
| Steinke & Vraciu 1999 | 9 | 18 | 2 | 7% | 31% |
| TOTALS | 108 | 835 | 175 | 16% (AVERAGE) | 10% (AVERAGE) |

aspects of linguistic semantics, e.g., formal aspects of meaning such as truth-conditional semantics, wherein one considers the conditions under which propositions are true or false and the consequences of recognizing such conditions. We do not see how such nonlexical semantics could be shared due to language contact.[3] Insofar as lexical elements are discourse-related and thus contribute to the pragmatics of the interpretation of utterances, however, they are subject to contact-induced change and so are treated here.

Although the lexicon is the unifying theme in this chapter, the result is eclectic for three reasons. The first stems from our view that one must consider the line between grammar and lexicon to be a fine or even indistinct one[4] and that this applies to certain Balkan phenomena.

A case in point is the expression of 'whether VERB or not' by means of verbal repetition wrapped around the negative marker,[5] i.e., VERB-'not'-VERB

3 Thus parallels between languages in the Balkans in aspects of their formal semantics, e.g., truth conditions or quantifier scope, are likely to be due to universality and not historical contact, except insofar as they pertain to the properties of specific lexical items. See the discussion on the interpretation of agents ("control") in gerundive constructions (§7.6.2.3.1.2) and on tense and aspect in subordinate clauses (§7.6.2.1.3.2–3) for some instances of nonlexical Balkan semantics. While these can be treated in formal semantics, any contact-relevant phenomena are essentially lexical in nature, associated with a particular grammatical formation and the morphemes constituting it.

4 In this way, we follow Pullum & Zwicky 1988 who say (p. 260) that the "lexicon is the repository of unpredictable phonological information about words" but add, nonetheless, that "it does not … contain *only* unpredictable information," allowing for certain regularities to be stated in the lexicon. Even Leonard Bloomfield, who famously in his 1933 work *Language* saw the lexicon as "a list of basic irregularities" (p. 274), recognized that existing patterns, whether compounds or syntactic strings, could be models for the creation of new forms (pp. 275–276), in a way therefore presaging the later notion of *lexical rule* (and see Aronoff 1988 on different senses of the term *lexical*, and Dixon & Aikhenvald 2002 on *word*).

5 This discussion is adapted from Joseph 2000a; on other types of repetition of elements in the Balkans, see §4.3.7 (on reduplication). The negation marker is the indicative negator (see §7.6.1), even in those languages (Greek, Romani, and Albanian) that distinguish indicative negation from modal negation (see §7.6, as well as §7.6.1 and §7.6.2), though Albanian does have a variant with the modal negator *mos* as noted below after example 4.1.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(Sandfeld & Olsen 1960: 47; Domi 1975; Banfi 1985: 79; Buchholz & Fiedler 1987: 506):[6]

(4.1) a. φύγει δεν φύγει 'whether one leaves or not' (Grk)
b. peniš se ne peniš se, šte te jam 'whether you foam or not, I'll eat you' (Blg)
c. spune nu spune 'whether he says (so) or not' (Rmn)
d. vjen s'vjen aq më bën 'whether he comes or not, I don't care' (Alb)
e. ladž na ladž o Roma vakerena peske (Jusuf 1974) 'Shameful or not, people are talking' (Rmi, Arli)

While this pattern is relatively productive in colloquial registers in each of these languages, there is one specific token that is shared more broadly across all the languages, namely with 'want', e.g., Grk θέλει δεν θέλει, Blg *šte ne šte*, Mac *saka nejkje*, Rmn *vrea nu vrea*, Alb *do s'do* (= *domosdo*), Aro *cu/di vreare, cu/di nivreare*, Rmi (Topaanli, etc.) *mangeja, na mangeja*, Jud *kyere, no kyere*, Trk *ister istemez*, all literally 'wants not wants' with the basic meaning 'like it or not.' This particular instantiation of the 'whether VERB or not' expression is likely to have been the starting point for the more general pattern in the Balkans, as indicated by a few key facts.

While the distributional evidence alone points toward this expression as the prototype within the Balkans, the Turkish expression *ister istemez* is especially significant. In Turkish-internal terms, it presents an irregularity: Turkish grammatically fixed verbal repetitions of the type VERB-GPRS.3SG VERB-NEG.GPRS.3SG normally mean 'as soon as. . .,' e.g., *gel-ir gel-mez* 'as soon as s/he comes' (lit., 'comes, doesn't come').[7] This fact suggests that *ister istemez* in Turkish could be a borrowing, as borrowings often stand out as irregular in some way synchronically.[8] Furthermore, there are non-Balkan parallels to 'want-not-want,' especially English *willy-nilly*, Latin *velit nolit(ve)*, though not so much for other specific tokens of the pattern, making 'want' seem like a particularly natural candidate for occurring in such a formation. As such, it would be a good starting point to consider for the entry of such a construction in any language.

One can go even further and suggest that while the Latin expression is a natural source to think of for the Balkan Romance instantiation of this formation, Greek could also be a possible source of the basis for this pattern in Balkan Slavic, if not the other languages, too. The reason for this assessment is that there is a prototype attested in early Postclassical Greek (Arrianus 3.9.16, second century CE), in the

6 And in some of the languages, also with other parts of speech, e.g., Rmi (Burgudži) *cikoro na cikoro, ama pišmani ka oves* (Jusuf 1974), 'whether he's [too] young or not, you'll be sorry' (lit., 'small.M NEG small.M but regret FUT become.2SG'). See §6.2.2.3.3 footnote 245 regarding the context of (4.1b).

7 This usage is calqued into Judezmo, e.g., *Lo vido no lo vido* (Varol Bornes 2008: 259) 'as soon as she saw him' (lit., 'him saw.3SG NEG saw.3SG').

8 However, there are a few other noncanonical meanings associated with the VERB-not-VERB expression in Turkish, and while these are not exactly 'whether or not,' they are closer to it than to the 'as soon as' meaning, e.g., *olur olmaz* 'ordinary, chosen at random' (lit., 'be be.not'), *bilir bilmez* 'half-wittedly, with little knowledge' (lit., 'know know.not') (Lewis 1967: 182).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

form θέλει οὐκ θέλει 'whether he wants to or not,' where οὐκ is the indicative negative marker that was current at that time, giving another early source for the pattern in the Balkans. On the other hand, the presence of this pattern in non-Balkan Slavic, e.g., Russ *hoćeš' ne hoćeš'*, raises questions either of its typological ("universal") likelihood or its spread as a learnedism.[9]

This example challenges the line between grammar and lexicon at the point where there was a single token of this type, that involving 'want,' as in Turkish. At such a point, one would be inclined to treat the formation as being lexical in nature, given that it is restricted to just one verb. Yet it seems that this single token was the likely basis for the creation of other parallel tokens, by an analogical extension of the model it offered. At some point, the several tokens of VERB-'not'-VERB, by clustering together, would be treated more economically by the recognition of a pattern in the grammar. However, questions such as how many such tokens are needed or whether those in the lexicon remain in the lexicon after the establishment of a productive pattern cannot be answered readily in a nonarbitrary (nontheory-bound) way. The boundary between grammar and lexicon is thus arguably a fuzzy one.

Therefore, in our discussion of the lexicon, some attention must be paid to phenomena that involve more than just individual lexical items; in some instances, patterns that show some, albeit limited, productivity as well as phrases that have idiosyncratic meanings and uses are considered. This expanded view of the lexicon guarantees a degree of eclecticism in any treatment of the lexical side of the Balkan sprachbund, but, as indicated above, there are two additional reasons for eclecticism here.

The second reason stems from the fact that the vocabulary of any language is never a closed set and will always contain items that reflect speakers' ability to converse with others on any topic. If one surveys the lexical stock of a language, there will be words that pertain to physical and intellectual culture, to different sectors of human endeavor, to the way humans interact with the natural world (including onomatopoeia), to the range of ways in which humans interact with one another – e.g., intimate, jocular, abusive, conversational, informational, ritual – and to any phenomenon susceptible to linguistic expression. As a consequence, the lexicon necessarily ranges over a wide array of meanings and real-world referents, and any discussion of the content of the lexicon necessarily presents an enormous range of potentially relevant tokens and concepts. Moreover, since we accept the view that no part of a language is exempt from the possibility of transmission through contact (see §3.2.1.7), all sectors of vocabulary are possible material for contact-induced transfer across languages.[10] The resulting study requires a consideration of a wide range of different types and classes of lexical items,

9 Culture areas of learnèd origin, e.g., the Indosphere, the Sinosphere, the Islamicate region, the region of Hellenicity in the ancient world, can and often do include sprachbunds, but these former should not be confused or conflated with the latter. While the former may include an oral component, they need not, and such a component will, by its very nature, not be characterized by social depth, i.e., it will be the property of elites. In contrast, the sprachbund is the result of both widespread and socially deep multilingualism, i.e., colloquial practice that is not limited to elites.

10 See §4.3 for some discussion of attempts to quantify borrowability, some of which attempt to demonstrate statistically that certain areas of vocabulary are systematically excluded from being

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

lexicalized phrases, lexically derived patterns, and the semantics associated with all of these.

The third reason for our eclecticism is chronological: since the Balkans have been a major contact zone for millennia, it is important to recognize different chronological layers of loanwords. However, although the discussion here is temporally eclectic, our primary focus is on the formative period for the sprachbund in the medieval and early modern periods, especially the Ottoman Balkans (see §1.1).

Nonetheless, we do draw some lines within this broad view of the lexicon. In particular, morphology per se, that is, the part of grammar that pertains to the form that words take in actual use, clearly has lexical ramifications. At issue are the derivation of new words – traditional *derivational morphology* or *word formation* – and the addition of inflectional material to stems in order to mark their relation to other elements in a sentence. Material that is clearly inflectional, i.e., with some relevance to syntax, is treated systematically in Chapter 6. However, some derivational material also has grammatical relevance, e.g., suffixes that derive nominals from verbs such that consequences for argument structure need to be taken into account. Such quasi-grammatical material is noted here but treated more fully in Chapter 6. Only more concrete types of derivation, processes that add truly lexical meaning to a stem, e.g., agentive-deriving affixes, are treated here.[11]

In what follows we therefore consider the lexical side to the Balkan languages in a selective way. There is a huge amount of material that is commonly discussed in treatments of the Balkan languages to which we devote only basic coverage, and at the same time there is one particular group of loans that we introduce here that does occupy our attention significantly. We defer detailed discussion of that type to §4.3, but in essence it comprises loans that are closely tied to conversational interactions, and we refer to these as *ERIC* loans, i.e., those that are "Essentially Rooted In Conversation." These conversationally based loans contrast with the loans that are more connected to aspects of material culture. It is this latter type of loan that has commanded the greatest attention of scholars over the years, but it is the former loans, the conversationally based ones, that in our view are more essential to understanding the nature of the formative processes behind the emergence of a convergence area, i.e., a sprachbund, in the Balkans. The bulk of this chapter, therefore, is devoted to motivating this conversation-based loan type and providing a substantial presentation of numerous relevant subtypes. In keeping with our interest in surveying material relevant to a full understanding of the historical interrelations among the Balkan languages, however, we also provide a brief, chronologically based overview of the various layers of loans in these languages without

borrowed. We reject any such claim, given the evidence that the right social situation can permit any type of word to be borrowed. See also footnotes 86 and 87 below.

11 In this way, we are deliberately bypassing theories of word-formation that put inflectional and more grammatically oriented derivation into the lexicon as part of the lexical listings of whole word forms; our goal is not to score points about morphological theory but rather to present the facts about the Balkans in a coherent, accessible, and theory-neutral way.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

attempting exhaustive coverage. The reader is directed to any of the standard handbooks mentioned above and specific works referenced below for further discussion of such contentful loans. The result is an overview of the full scope of relevant lexical material.

In discussing Balkan lexis, we must address the issue of register. Since many of the forms we focus on are by their very nature colloquial, normative judgments about their use are basically irrelevant, except insofar as such evaluations come to affect spoken usage. At the same time, however, questions of spatial and temporal distribution, i.e., whether or not a particular lexical item is *dialectal* in the sense of restricted to certain dialects or *obsolete/historical* in the sense of restricted to earlier time periods, must be addressed. Issues of social distribution, i.e., whether an item is felt to be pejorative, vulgar, technical, limited to professional jargon, etc., can also be relevant. For our purposes, however, that which is of greatest interest is the movement of vocabulary from one language to another. To the extent that this movement is in some way spatially, temporally, or socially limited, and that such limitation is of immediate relevance, we note it. In general, however, we do not attempt to classify each individual item according to whether or not such limitations hold. Those are the concerns of dictionaries and specialized studies.[12] Our primary focus is the general fact of occurrence itself.

## 4.2 Overview of Commonly Discussed Material

Besides the various colloquial lexical effects of contact to be documented in §4.3, the vocabularies of the respective Balkan languages have been augmented by the entry of numerous, mainly contentful, foreign words associated with specialized lexical domains or various sorts of cultural or social contexts. For the most part, these content loans are tied to different historical phases, and no discussion of the Balkan lexicon would be complete without some consideration of them. At the same time, however, these more content-bound borrowings are arguably less indicative of a sprachbund than ERIC loans. The conversational loans are the result of sprachbund-conducive conditions, whereas the other, more contentful loans are not as distinctive by themselves. While traditional content loans have long been taken as characteristic of the Balkan sprachbund, it is in fact the conversational loans that are diagnostic. While the presence of the conversational loan-type presupposes the presence of some nonconversationally based loans, the reverse is not the case. In other words, ERIC loans are a vital component in Balkan linguistics, while these others relate more to the linguistics of the Balkans.

Accordingly, we discuss these contentful loans without attempting an exhaustive coverage of various bilateral and other localized and temporally disparate contact

12 See, for example, Sobolev 2004 on the relative distributions of various Turkisms in selected Balkan dialects, which includes lexicographic material, and Leschber 2007 on Turkisms in current youth slang. Slang and other highly variable registers constitute, in a sense, a moving target. To some extent, therefore, our treatment must be taken as a snapshot.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

situations that have yielded limited lexical influences in a larger Balkan context. Rather we take a basically chronological approach from ancient times to the present.[13] Thus, for example the lexical effects of Greek on Arvanitika or on the Romani of Ágios Athanásios (older Ali Bey Köy, now a suburb of Sérres), as opposed to that on the Romani of Agía Varvára (a suburb of Athens), whose speakers arrived from Turkey with the 1923 exchange of populations, or that of Balkan Slavic on Sarakatsani Greek in northeastern Greece and southern Bulgaria, are outside the focus of this chapter unless such contact shows some interesting development or has consequences in other domains, as in the case of ERIC loans or as with the reverse phonological interference in Arvanitika due to contact with Greek (see §5.2).[14]

### 4.2.1 Borrowing of Content Words – Historically Identifiable Layers of Vocabulary

As noted previously, there is an enormous literature on various types of content words in the histories of individual Balkan language contact situations. We can cite here some of the more important, representative studies. Desnickaja 1963, Svane 1992, Ylli 1997–2000, and Omari 2012 cover Slavic loans in Albanian and Capidan 1925a examines Slavic in Aromanian; the study of the influence of Slavic on Romanian goes back to Miklosich 1862 and Leschber 2012 is a recent contribution; Tietze 1957 treats Slavic loans into Turkish, and Vasmer 1941 is the classic source on Slavic toponymy in Greece, and Meyer 1894 examines Slavic, Romance, and Albanian loans into Greek; Boretzky 2012 offers a comprehensive survey of the Greek lexical influence on Romani; Tietze 1955, 1983; Symeonidis 1973, 1976; and Tzitzilis 1987 discuss Greek loans into Turkish; Tzitzilis [Dzidzilis] 1990 examines Greek loans in Bulgarian; Jašar-Nasteva 1953abc surveys Albanian loans into Macedonian; Vrabie 2000: 71–84 devotes considerable space to the sources of the Aromanian lexicon, including loans from Albanian, Slavic, Greek, and Turkish; and Meyer 1888a, Jokl 1936, Haarman 1972, and Bonnet 1998 consider the Latin element in Albanian. Haarman 1978 examines the Latin element in the Balkan languages in general; Paşcu 1924 surveys Romanian elements in the Balkan standard languages of that period: Albanian, Bulgarian, Greek, Bosnian-Croatian-Montenegrin-Serbian (BCMS), and Turkish. Stankiewicz 1964 discusses loanwords and derivational affixes of Balkan and Slavic origin in the Judezmo of former Yugoslavia. Also, Bunis 2017 makes the point that while Slavic loans in Judezmo were relatively rare – especially in comparison with Turkisms (cf. Yenisoy 2015) – until the

13 Sandfeld 1930, by contrast, used an organizational schema based only on the source language, itself a problematic concept when the time span is millennial.

14 For Greek on Arvanitika see Tsitsipis 1998, Sasse 1991, and Hamp 2007; for the Romani of Ágios Athanásios see Sechidou 2011; and for the Romani of Agía Varvára see Igla 1996. Hoeg 1925–1926 is still the best general account of Sarakatsani Greek, but see also Skok 1927, Tzitzilis 1999, and Kahl 2007.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

nineteenth century, as Slavic nation-states acquired independence and their respective national languages developed prestige, the number of Slavic loanwords in Judezmo increased. Cazés 1999 and Bunis 1982 are also relevant here. On Judezmo influence on Turkish, see Rocchi 2007.

The influence of Turkish on the various Balkan languages has been of major interest since Miklosich 1884–1890, and today there are specialized dictionaries and studies of Turkisms for almost all the Balkan languages: Boretzky 1975–1976; Dizdari 2005; Latifi 2006, 2012, 2015; Lloshi 2020; Bufli & Rocchi 2021; Lleshi & Rugova 2023 for Albanian; Polenakovikj 2007 for Aromanian; Grannes 1970; Grannes et al. 2002 for Bulgarian; Georgiadis 1974; Kukkidis 1960; Kyranoudis 2007; and Orfanos 2014 (also, de facto, the list given in Dizikirikis 1975, discussed in §4.4, and the index in Georgiadis 1974) for Greek; Jašar-Nasteva 2001 and Cvetkovski 2017 for Macedonian; Șaineanu 1900; Wendt 1960; Drimba 1992–1993, 2001; Altay 1996; Suciu 2010 for Romanian; and Knežević 1962 and Škaljić 1966 for BCMS. We can also note here Graham 2020, which provides unique insight into the adoption of Turkisms by non-Muslims in early modern Bosnia and Bulgaria. For Romani there is Friedman 1989c, which is based on Messing 1988. For Judezmo, Bunis 1999, 2023, and Dobreva 2016 provide relevant discussion, but see also Danon 1903, 1904, and 1913. Finally, we can mention two dictionaries that compare Turkisms across various Balkan languages, Rollet 1996 and Karaağaç 2008, which latter has a much broader range in terms of languages, bibliography, and vocabulary.

Romani lexical elements are to be found in all the Balkan languages, but Leschber 1995 and Bochmann 1999 are among the few works devoted to the study of such elements, and both examine Romanian (see also §4.4.2 and footnote 372). Finally, Latifi 2015 gives a comparative overview of Balkan Turkisms. We note also in this regard Asenova & Detrez 2021, which is limited to common loanwords in Albanian, Bulgarian, Greek, and Romanian (with some references to the former Serbo-Croatian) but is a useful source of shared vocabulary for those languages and gives etymological information (using Bulgarian headwords) for the various loans, mostly Turkish, but also Slavic, Greek, and Latin. It does not include the Albano-Romanian commonalities as the work is limited to words occurring in at least three out of the four languages studied. Owing to the book's linguistic limitations, various commonalities that could be seen if other languages with literary standards had been considered are absent. Thus references to Aromanian, Macedonian, and Romani, all of which have literary standards, are ignored in this work. A case in point of the resulting lacunae is Albanian *bardhë* 'white,' which was borrowed into both Aromanian and Macedonian (see discussion in §5.3). Still, within its limitations, this work is useful.

The reader is referred to these works for discussion of details that cannot be treated here, our focus being only the broad outlines of contributions to various historically identifiable strata of lexical borrowing across the Balkans.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

#### 4.2.1.1 Non-Greek Paleo-Balkan Vocabulary

There is a layer of vocabulary in various Balkan languages that is identifiable as very old but not Greek (the oldest well-attested language in the Balkans). Under this rubric, there are two types of words that deserve mention: (1) those shared by Albanian, Romanian, and other Balkan languages and (2) those shared only by Albanian and Romanian. We refer to them here as "non-Greek Paleo-Balkan" in an attempt to characterize them as neutrally as possible. They are ancient, but they cannot be dated. Some of them are definitely Indo-European, and some are undoubtedly from the language that is ancestral to Albanian (cf. Hamp 2007: 373–395), but not all can be definitively assigned to any specific ancient non-Greek Balkan language. Some words may be pre-Indo-European. These problems have not prevented speculations regarding specific ancient languages, but we eschew such issues, as they do not bear directly on our purposes here. These words are thus "Paleo-Balkan" and "non-Greek." Taking them to be old has led them to be assigned to a so-called Balkan substratum, the existence of which is a reasonable assumption, but still it is a concept for which adequate linguistic details are lacking, other than it probably having an Indo-European component itself (see §1.2.1 for related discussion).

Much of the pan-Balkan old layer consists of words that are associated with pastoral life, animal husbandry, and various domestic items and activities (see also §1.2.3.1). The number of such words is around twenty, depending on the judgment of various linguists. (See Neroznak 1978, especially pp. 186–216, and sources therein; for the ancient Balkan languages, Katičić 1976 remains authoritative, though see also Woodard 2004; Hamp 2007 is also an important source. See also Sobolev 2003: 332–349 and Borescu 2018.) We note just an illustrative sampling of forms from various languages, leading with Albanian for purely alphabetical reasons:

(4.2) Alb *balgë/bajgë/bagël* 'animal manure,' BCMS *balaga/balega* 'excrement,' BRo *baligă/balegă* 'droppings', Mac (dial) *balega* 'manure'

Alb *drugë*, Aro *drugă*, Grk ντρούγα/δρούγα, BSl (dialectal) *drug*, BCMS *druga* 'wooden bobbin, distaff'

Alb *shtrungë,* Aro *strungã*, Rmn *strungă*, Mac *straga*, Blg *străga/stărga* 'enclosure or narrow passage for milking sheep or goats, separating lambs/kids, etc.'; also Mac and WBlg *strunga* 'idem' (from Aro), Grk (Epirus and Sarakatsan) στρούγγα 'dairy'(see Hamp 1977b)

Alb *vatër* (DEF *vatra,* Geg *votër/votra*) 'hearth, fireplace,' BRo *vatră* 'hearth, etc.,' BSl (dialectal, esp. Mac and WBlg) *vatra* 'hearth,' BCMS *vatra* 'fire,' Grk (Sarakatsan) βάτρα 'fire, flame, hearth' (see Hamp 1976, 1981)

As with the next group, there is considerable controversy regarding the precise sources for these words, both within and outside the Balkans. Thus, for example, *vatra* occurs as 'fire' in Ukrainian, 'hearth, fire, dying ashes' in Polish (orthographic *watra*) and Czech (East Moravia), 'camp fire' in Slovak (also southern Poland and part of Czech), and 'poker' in Slovene (see Udler 2000 and Hamp 2007: 373–382, who argues that *vatra* derives from the ancestor of Albanian and spread

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

via dialect chains). Regardless of the precise age and provenance of these loans, however, there is not much more to say about them here other than that they do point to ancient language contact.

The other group of words is thematically broader (although still arguably limited to a notional concept of basic vocabulary) but restricted to just Albanian and Balkan Romance,[15] and thus, as noted briefly in §1.2.1.4, they are taken to be old shared vocabulary that link the ancestor of Albanian, on the one hand, and the language whose speakers shifted to the Latin that became Balkan Romance, on the other. There are perhaps approximately seventy such words, although there is some disagreement concerning how many of these words belong here (see Polák 1958; Kalužskaja 1977; Ismajli 2015: 271–467, and sources cited therein). Some belong to pastoral vocabulary, thus overlapping semantically with the first group, but their restriction to these two languages is taken to be significant. Several of these have clear Indo-European sources, e.g., Alb *sorrë* – Rmn *cioară* 'blackbird' (from PIE **kwērsnā*), and some IE words seem not to be directly inherited from an Italic genealogical predecessor.[16] For instance, Rmn *druete* 'woods' and Alb *dru* 'wood' clearly derive from the PIE **deru-* 'wood' (most likely in a zero-grade form **dru-*), seen in Eng *tree*, AGrk δρῦς 'tree' and δόρυ 'spear', etc., but forms of this word are absent from Latin or any other ancient Italic language, so that the appearance of a derivative of this PIE stem in Balkan Romance is unlikely to be the result of a direct inheritance from Latin; hence, it is judged as a substrate word. A sampling of others is given in (4.3):

(4.3)

| Romanian | Albanian |
|---|---|
| bucurie 'joy' | bukuri 'beauty' |
| buză 'lip' | buzë 'lip' |
| ceafă 'neck' | qafë 'neck' |
| ciump 'end, snag' | thumb 'tack, stinger' |
| coacăză 'currant' | kokë(z[ë]) '(little) head [blackhead, a poultry disease]' |
| mal 'riverbank' | mal 'mountain' |
| moş 'old (man)' | moshë 'age' |
| mugur 'bud' | mugull 'bud' |
| ţap 'billy goat' | c[j]ap ' billy goat'[17] |

15 The most interesting and important matches here involve Romanian, rather than SDBR, however, since Aromanian, and Meglenoromanian matches with Albanian could in some instances be later borrowings. Hence our citation of Romanian forms.

16 The root for 'black' (OCS *črьnъ*, Skt *kṛṣṇá-*) is ambiguous between initial **k* or **k$^{w}$*; however, in order to connect it with *sorrë*, an attractive match on semantic grounds, Hamp 2007: 323–325 has posited **k$^{w}$*, a reconstruction which works well for the Slavic and Sanskrit forms, **k* being contraindicated since one would expect **kē* to yield Albanian [ko] as in *kohë* 'time,' if cognate with OCS *časъ* 'hour' from **k(*$^{w}$*)eH$_1$s-* (cf. Derksen 2008: 79) or **k(*$^{w}$*)eH$_1$sḱ-* (Hamp). Others connect *sorrë* with the *kor-* of AGrk κόραξ, Lat *corvus* 'raven,' or the *svor-* of Proto-Slavic **svorka* (OCS *svraka*) 'magpie' (see Meyer 1891: 390), though those connections entail other phonological problems that make them unattractive in our view.

17 The Albanian *th-* in *thumb*, is presumably from a PIE **ḱ*, whence the Romanian *c(i)-*. The Balkan and extra-Balkan distribution of 'billy goat' illustrates the problems of differentiating the two types of vocabulary discussed in this section. In BCMS, this word only occurs in the extreme northwest (Istria, Čakavian, Slovene), although the pejorative meaning 'scraggly beard' in the BCMS of Kosovo is undoubtedly also from this word. In Greek, τσάπος is limited to Epirus. *Cap* also occurs in Hungarian, Polish, and Ukrainian, where it is presumed to be of Romanian origin. While the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

As noted in §1.2.1.4, the exact nature of the prehistoric link that these indicate between the languages is extremely controversial;[18] see also §6.1.2.2.1.3 (Hamp 1982) and §7.9.2 for discussion of some possible morphosyntactic and syntactic matchings between Albanian and Romanian that might be old in the same way as these lexical matchings.

### 4.2.1.2 Latinity (The Roman Era)

Romans settled the Balkans definitively in the second century BCE with the conquest of the Illyrians and the Macedonian kingdom. In subsequent years, they spread their influence further in the general region. As noted in §1.2.3.3, there is an important historio-geographic construct in the Balkans that has a key linguistic correlate that helps to define aspects of the Roman era. This is the *Jireček line* (or *Jireček-Skok line*), running west-east from the Adriatic to the Black Sea across modern central Albania, northern Macedonia, and central Bulgaria then north along the coast to Dobrudja, which demarcates the respective extent of Roman and of Greek influence by reference to the predominant language of inscriptions: north of the line it is mostly Latin and south of the line mostly Greek, with an area of bilingualism between Jireček's and Skok's demarcations (see §1.2.3.3). This is not to say that Latin was unknown in the southern Balkans, or that Greek was unknown in the more northerly regions, but it gives an idea of the administrative reach of Latin in the Balkans in the early Christian period.

Not surprisingly, there is a large Latin lexical presence in the entire Balkans. To get an idea of the size of this presence, consider these numbers compiled by Mihăescu 1978: 30ff. in his discussion of words from Latin in the Balkan languages. He notes, for instance, that there are some 3,000 terms of Latin origin in Byzantine (Greek) literature, of which 207 survive into the modern language. When viewed by semantic category, these numbers break down as in Table 4.2.

These items are most concerned with dimensions of public life under the control of Roman governance. The penetration of these words into Balkan languages took place over a period of almost a millennium, and it was particularly intense in the fourth through sixth centuries CE. Many of these words are found as well in Albanian, as loanwords, and in Balkan Romance, as inherited items or via secondary spread, and since the influence lasted into the time of the entry of the Slavs into the area, Latin words are found as well in Balkan Slavic.[19]

Epirus and Kosovo forms could both be later borrowings from Albanian, the northwestern BCMS suggests the possibility that the word belongs specifically with the Albanoid-pre-Romanian lexicon (Skok 1971: 251).

18 To the references mentioned there, we add the various works by Eric Hamp over the years, many of which are to be found in Hamp 2007.

19 As Sandfeld 1930: 54 observes, a real problem for Balkan Slavic is determining whether a given loanword entered directly from Latin or via an intermediary such as Gothic; see also footnote 20. Thus, for example OCS *vino* comes ultimately from Latin *vīnum*, but whether directly or via Gothic is disputed. The final *-o* points to a Gothic intermediary unless one makes other assumptions, which, while possible, are unprovable given our current state of knowledge (cf. Skok 1972: s.v.). See also

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.2 *Latin loans into Greek by semantic domain*

| | In Byzantine Greek | In Modern Greek |
|---|---|---|
| Military terms: | 431 | 30 |
| Public administration: | 384 | 14 |
| Judicial activity | 341 | 2 |
| Clothing | 145 | 12 |
| Religion (plus Calendar) | 118 | 18 |
| Flora | 178 | 1 |
| Weights & Measures | 76 | 10 |
| Fauna | 73 | 9 |
| Court life | 67 | 9 |
| Other | 1099 | 103 |

A few illustrative examples (4.4) provide instances of different types of words of Latin origin in the Balkans. They are presented here merely as representative; readers are referred to sources such as Mihăescu 1978: 30ff., Skok 1928, and Sobolev 2003: 270–289 for more examples:[20]

(4.4) Lat *acetum* 'vinegar': BSl *ocet* (OCS *осьтъ*), Alb *uthull* (cf. *oftull* in Pulevski 1875: 93) Rmn *oțet* (< BSl)
Lat *būbalus* (VLat **būvalu*) 'water buffalo': BSl *bivol* 'idem' from BSl into Rom *bivol* (but Aro *buval*) cf. Grk βούβαλος, βούβαλις, BER I: s.v.)
Lat *camisia* 'shirt': MedGrk καμίσι, ModGrk πουκάμισο (with που- from υπο- 'under') 'shirt,' Alb *këmishë* 'shirt,' Rmn *cămaşă* 'shirt'
VLat *coctorium* 'oven': Rmn *cuptor* 'oven, furnace, kiln,' Blg *kuptor* (from Rmn), Alb *koftor* 'pot-bellied heating stove'
Lat *centum* '100': Alb *qind* '100'
Lat *fossatum* 'military trench': MedGrk φουσάτο 'army,' Alb *fshat* 'village' (originally 'fortified settlement'), Rmn *sat* 'village' (ditto)
Lat *furca* 'fork': Rmn *furcă* 'pitchfork,' BSl *furka* 'spindle,' Alb *furkë* 'pitchfork, spindle,' MedGrk φούρκα 'gallows'
VLat **furnu* 'oven': Med/ModGrk φούρνος, Aro *furnu*, BSl *furna*, Alb *furrë*, Trk *furun*
Lat *hospitium*: Med/ModGrk σπίτι 'house,' StAlb *shtëpi* (Geg *shpi*) 'house'
Lat *imperator*: Alb *mbret* 'king'
Lat *lucta* 'struggle, fight,' Alb *luftë* 'war,' Rmn *luptă* 'war'
Lat *paganus* 'peasant (later 'pagan')': OCS *poganin* 'pagan, evildoer' (Mac *paganin* 'pagan' Blg *pogan* 'unclean'), Alb *pëgërë* 'unclean, dirty,' *pagan* 'pagan,' Rmn *păgîn* 'pagan,' Aro *pîngîn* 'pagan' (see Duridanov 1999: s.v. for details)
Lat *pomum* 'apple': Alb *pemë* 'tree,' Rmn *pom* 'tree'

Atanasov 1993 and Pronk-Tiethoff 2013: s.v. On the other hand, the /i/ from Latin /ū/ and /o/ from short /a/ and loss of the final /u/ is exactly what would be expected in an ancient loan.

20 Some words that are ultimately of Latin origin entered Slavic via a Germanic intermediary, e.g., BSl *oltar* 'altar' (OCS *олътарь*) is from OHG *altāri*, which in turn is from Latin *altāre* 'idem' (BER IV: s.v.). See now Pronk-Tiethoff 2013 on the problem of old Germanic loans in Slavic, but also Lunt 1982. Albanian *altar* is directly from Latin.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Lat *rosalija* 'early summer festival when graves were decorated with roses': OCS *rusalija* 'Pentecost'
Lat *sagitta* 'arrow': MedGrk σαΐτα 'dart,' Alb *shëgjetë* 'arrow,' Rmn *săgeată* 'arrow'

And, as discussed in greater detail below in §4.3.1.8, late Latin is the ultimate source of two terms that are quite outside of these lexical domains, in an entirely different semantic sphere, namely that of the family: the wide-ranging cluster of related forms exemplified by ModGrk νονός 'godfather,' Mac *nunka* 'godmother,' BSl *kum* 'godfather, best man etc.,' Alb *kumbarë* 'godfather,' ModGrk κουμπάρος 'best man.'

### 4.2.1.3 Greek in the Balkans

Greek words that must be of ancient provenance are to be found in some of the Balkan languages. In particular, Albanian has, for instance, *drapër* 'sickle' (Geg *drapën*) from Ancient Greek δρέπανον, *lakër* 'cabbage' (Geg *lakën*) from λάχανον, *mokërë* 'millstone' (Geg *mokënë*) from (Doric) μᾱχανᾱ́, and *tarogzë* 'helmet' from θωρᾱ́κιον 'breastplate, armor,' where, e.g., the *t-* of this last form points to its great antiquity, from at least before the Hellenistic Greek shift of ancient < θ > from [$t^h$] to [θ] (see Horrocks 2010: 170–171; Jokl 1984). And there are later loans that are still somewhat old, such as Albanian *fnazë* 'light fall of snow' (Newmark 1998: 231) from νιφάδιον 'snowflake,' where a post-Classical date is suggested by the *f-* for ancient < φ >.

There are numerous loans from Greek into Old Church Slavonic that fall into many of the same classes as Latin loans discussed in §4.2.1.2; Vasmer 1907 shows that the categories of Greek loans include names of plants, animals, minerals, humans, body parts, nature, and home-related items, and that they fall into three distinct chronological phases: (1) early borrowings into Common Slavic, e.g., BSl *koliba*, 'hut, cabin' (whence Rmn *colibă*, Trk *koliba*, Alb *kolibë*, cf. Aro *cãlivã*), Grk καλύβα, AGrk καλύβη (Vasmer 1907: 243; BER II: s.v.), where the realization of AGrk /a/ as CoSl /o/ in the first syllable points to a period before the reinterpretation of quantity as quality; (2) borrowings before the conversion of the South Slavs to Christianity, e.g., OCS *kǫponi*, Blg *kăponi*, Mac *kapan* 'scales [for weight]' from MedGrk καμπάνα (Vasmer 1907: 251; SSl *kambana* 'bell' is a later loan from the same source); and (3) those from the Christian era (see items in (4.5)). The emergence of Eastern Orthodox Christianity in the first millennium CE, with Greek as its liturgical language, together with the Christianization of the South Slavs in the ninth century and the subsequent spread of Orthodox Christianity in the linguistic Balkans, was in many ways a watershed period for the influence of Greek, as it led to the introduction of Greek ecclesiastical terms into various languages of the region, as seen in (4.5) (cf. Sandfeld 1930:

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

20–21; Sobolev 2003: 204–269, as well as Budziszewska 1969 on Greek loans in Bulgarian):[21]

(4.5) Grk ἁγίασμα 'sanctification': ChSl *agiazma*, Blg *agiazma/ajazma*, Mac *ajazma* 'holy water' Alb *ajazmë*, Rmn *aghiazmă*, Aro *(a)yeasmó* 'holy water'
Grk ἀναφορά 'blessed bread': OCS *(a)nafora*, BSl *nafora* 'holy or toasted bread' Alb *naforë*, BRo *(a)naforă*
Grk ἀνάθεμα 'curse, excommunication': BSl *anatema* (also Mac *natema go* 'damn him,' etc.), BRo *anatemă*, Alb *anatemë*[22]
Grk εἰκόνα 'icon': OCS *ikona*, BSl *ikona*, Alb *ikonë*, BRo *icoană*
Grk καλόγηρος 'monk': OCS *kalogerъ*, Blg *kaluger*, Alb *kallogjër*, BRo *călugăr*
Grk ἡγούμενος 'abbot': OCS *igumenъ*, Blg *igumen*, Mac *egumen*, Alb *(i)gumen*, BRo *egumen* (*igumen*)

There are also some borrowings from Greek of a more grammatical, nature, such as: Alb *anamesa*, Aro *anamasa* 'in the middle' from Grk ἀνάμεσα, Aro *anda* 'when' from Grk ὅντα, Alb *andis* 'instead of' from Grk ἀντίς, BSl *oti* 'that' from Grk ὅτι; see §§4.3.3.2, 4.3.3.4 for more details on such loans. Moreover, although many loans from Greek into Balkan Romance, especially Romanian, passed through Slavic (see §4.2.1.4), there are some that are found in all of Balkan Romance without occurring in Slavic, suggesting an early point of entry into Vulgar Latin from Greek, most likely soon after Latin entered the Balkans; Sandfeld 1930: 30 notes, for instance, Rmn *proaspăt*, Megl *proaspăt*, Aro *proaspit* 'fresh,' from Grk πρόσφατος 'recent' (cf. also Nevaci 2015). Tzitzilis 2001b discusses loans from Medieval Greek into Romani, e.g., *kurko* 'Sunday' from MedGrk κυρικόν (ἦμαρ) 'idem.'

### 4.2.1.4 Slavic

As Slavic speakers entered the Balkans in the sixth to seventh centuries, they began to exert lexical influence in the region (see Sobolev 2003: 324–331 for some examples). Svane 1992 and Ylli 1997–2000 documented extensively, for instance, the hundreds of Slavic loanwords of various types and in various domains to be found in Albanian, including farming terms (e.g., *plug* 'plough,' *lopatë* 'shovel,' *oborr* 'yard'), foods (e.g., *kastravec* 'cucumber'), clothing items (e.g., *opingë* 'sandal'), flora and fauna (e.g., *sokol* 'falcon,' *ljubiçice* 'violet'), items pertaining to social order (e.g., *rob* 'slave'), and other cultural loans (e.g., *pusullë* 'note'). Based on the evidence of sound changes in the respective languages that the loans show, or fail to show, it appears that while some seem to be relatively early (c.700–1000CE), e.g., *porosit* 'order; request' (cf. Common Slavic **porõčiti* 'idem;

21 The influence of Greek on Slavic was so strong that early versions of Cyrillic included the Greek letters theta, ksi, and psi, exclusively for the representation of Greek words, and qoppa for the numeral '90.' Even the older Glagolitic alphabet included theta. (The representations of /o/ and /i/ in the two alphabets are complex and need not concern us here.) See §4.2.1.4 on Slavic as the intermediary for some Greek terms in Romanian.

22 Here the Albanian /t/ suggests a non-Greek intermediary.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

entrust') and some rather late (post-1500CE), e.g., *banak* 'counter' (cf. Srb *banak* (stem: *bank-*) 'shelf,' most seem to have entered between 1000 and 1500.[23]

With regard to Greek, apart from Slavic place names (Weigand 1928; Vasmer 1941), only a relatively small number of Slavic loanwords can be identified that are in general use in the standard language or are generally recognizable today (Andriotis 1983: s.vv.; Babiniotis 1998, LKN), e.g., ρούχα 'clothes' (Slv *ruho* 'rag'), ντόμπρος 'honorable; noble' (Slv *dobro* 'good, well'), τσαντίλα 'sack for straining cheese' (Slv *cedilo* 'strainer'), and τσάσκα 'cup' (Slv *čaška*). See, however Filipova-Bajrova 1970 for a number of other examples. Still others occur that are regionally more restricted in their distribution; Weigand 1928: 33 mentions τσέλιγκας 'shepherd,' found in the Heptanesia and Acarnania, from Slv *čelnik* 'leader of a clan; lead shepherd,' and Budziszewska 1991 has identified several hundred words of Slavic origin that are reported for various locales, mostly, but not exclusively, in northern parts of Greece, e.g., nouns like ζακόν 'custom' (Flórina, Kastoria, Grevena, etc.), from BSl *zakon*, and numerous words for flora and fauna, e.g., κλένος 'maple,' from BSl *klen*, γουστέρα 'large lizard' (Larissa, Lamia), from BSl *gušter(a)*, and κναβ (= κουνάβι, with northern high vowel loss) 'marten' (Thessaly, Thrace, etc.) from BSl *kuna.*

In the case of Balkan Romance, Slavic had a tremendous direct influence and also served as a conduit for many Greek words entering Romanian, especially those associated with Eastern Orthodoxy. Among the nonreligious terms that passed from Greek to Slavic to Romanian are *ieftin* 'cheap' (ChSl *jevtin*, Blg *evtin* [eftin], MedGrk εὐθηνός (ModGrk φθηνός), and *a mirosi* 'smell' (ChSl *mirosati*, Grk μυρόω, with aorist stem μυρωσ-).[24] Religious terms include those cited in §4.2.1.3, such as Rmn *icoană* 'icon' (cf. Grk εἰκόνα, OCS *ikona*, Blg *ikona*), or Rmn *călugăr* 'monk' (Grk καλόγηρος, ChSl *kalogerъ*, Blg *kaluger*). Church Slavonic was the language of literacy in Romania into the modern period, and the process of its being replaced by Romanian begins in documents of the sixteenth century. It took until 1863, however, for the Romanian Orthodox Church to officially change the liturgical language; this was around the same time that the Latin alphabet officially replaced Cyrillic, although the use of Cyrillic continued in Romania until the 1920s. Thus lexical influence from Slavic in the domain of religious terminology continued for longer than in other domains.

23 Precise dating is complicated in some instances by the possibility of the word being due to pre-Balkan contact, e.g., *baltë* 'mud' (B. Demiraj 1997: s.v.; see also Duridanov 1977: 691) vis-à-vis OCS *blato* 'swamp.' In the case of, e.g., *sundoj* 'rule' vis-à-vis OCS *sǫditi* 'judge,' the preservation of Slavic nasality could represent the conservatism of local Slavic dialects, which, in some parts of Albania, still have nasality preserved as homorganic nasal sonorants before stops (cf. Seliščev 1931: 291–292 and Friedman 2018a, and see §5.4.1.1iv).

24 The initial /j/ in Romanian indicates that the word entered during the period when initial /e/ in Slavic was automatically realized with a prothetic /j/ and before the loss of that prothetic /j/ in East South Slavic.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.2.1.5 Romance and the Crusades

The later Middle Ages, covering roughly the period from 1100–1453 in some reckonings (e.g., that of Browning 1983: 69ff.) and also beyond that into the early modern era, saw the Balkans and the languages of the Balkans come into contact with Western Europeans of the day as a result of the Crusades and subsequent events. The Fourth Crusade especially had linguistic consequences, as Constantinople was captured and sacked in 1204, and what is commonly referred to as a Latin Empire was in parts of former Byzantium. Writing about Greek, Browning 1983: 70–71 notes that "the effects of the Latin conquest were complex [as] Latin loanwords flooded into the language." He considers it important to clarify that "in this context 'Latin' refers not to the classical language of Rome, but to the Romance vernaculars spoken in the Mediterranean area." Thus "Latin" influence on the Balkan lexicon at this stage is actually Romance influence, though due to the somewhat diglossic relationship between medieval Latin and vernacular Romance in this era, the written language continued to exert some influence on the spoken. As the Romance varieties became more clearly distinct languages, it is possible to speak of influence from particular languages as one moves into early modern times, i.e., the sixteenth century and beyond, although specific Romance dialects can sometimes be identified as sources during earlier periods.

The most direct Romance influence was on those parts of the Balkans that came under the dominion of various Romance rulers. Thus, in addition to Western European control of Constantinople, by the end of the thirteenth century the Franks controlled much of the Peloponnesos, and other parts of Greece, including strategic harbors on the Peloponnesos, and the islands of Euboea, Crete, and some of the Cyclades came under the dominion of Italian states, primarily Venice but also in some instances Genoa.[25] There was also Venetian and central Italian control and influence along the coast of the Adriatic Sea and the Ionian Sea, thus affecting Albanian, and, even though it is a language area generally outside of the purview of this book, also along the Dalmatian coast, affecting BCMS. Some lexical influence on other languages can be seen, but it is indirect, mediated via Greek, Albanian, or BCMS. Moreover, a considerable amount of maritime and nautical terminology spread from Italian and Venetian all over the Mediterranean, and thus throughout the Balkan languages, including Turkish.

Among such "Romance Latinisms" are (Middle) Greek ἐξόμπλιον 'example' (cf. Frn *exemple*), μισὶρ 'sir' (cf. Frn *monsieur*), τσάμπρα 'room' (cf. Frn *chambre*), ῥόη ῥόι 'king' (cf. Frn *roi*), among many others.[26] Venetian influence, distinguishable by characteristic lexis and phonology, is seen in Grk βελούδο 'velvet' (Vtn *veludo*, vs. Itl *velluto*), τζογος 'gambling, card-playing' (Vtn *zogo*, vs. Itl *giuoco*), αϊδάρω 'help' (Vtn *aidar*, vs. Itl *aiutare*), and κουζίνα 'kitchen' (Vtn *cusina*, vs. Itl *cucina*). For Albanian, one can note *rrugë* 'road' (Itl/Vtn *ruga*), *kanal* 'canal' (Vtn

25 Cyprus, too, was Romance-controlled, although this is less relevant for the Balkans per se.

26 See Triandaphyllidis 1909 for details on external lexical influence on Medieval Greek; Browning 1983: 72ff. offers a brief but useful survey, as does Banfi 1985: 96ff.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

*canal*), *kasellë* 'storage chest' (Itl *cassella*), and *frat* 'brother in a monastic order, friar' (Itl *frate*), inter alia. Some of the Albanian forms are now nonstandard, and Italian forms especially entered Geg, since the historically Catholic population there was in direct contact with Italian Catholics, and northern Geg coastal regions were ruled by Venice. In some instances, phonological differences pointing to differential time of borrowing – and thus differential language source – can be observed; for example, beside the early Albanian borrowing *lter* 'altar' (Newmark 1998: 466) from (Late) Latin *altare*, there is also *altar* 'altar' from Itl *altare*.

Examples of the indirect influence, showing spread of Latinate/Romance loans from one of the directly affected languages into another language, also occur. For instance, Blg *pogáča,* Mac *pógača* 'round white loaf' are from BCMS *pògača/pogȁča*, but there are also Rmn *pogáce* (dialectal *bogáce*), Aro *pugace, puɣace*, Alb *pogaçë* (dialectal *pugaçë*) 'pogacha,' MedGrk μπογάτσα/μπουγάτσα (late MedGrk πογάτσα) 'a kind of dairy pie,' Trk *poğaça*, dialectal *boğaça* 'small round roll,' as well as Hung *pogacsa* and dialectal Grm *Pogatsche*, all ultimately from OItl *focācea = focaccia* 'cake' (Lat *focacius* '[bread cooked in the] fire/hearth'); see BER V: 421–422 for discussion of various possible routes. It is generally accepted that BCMS is the source for BSl, which in turn is suggested for Aro and Alb; the Romanian could be from BCMS or Bulgarian or even Hungarian, whereas the Greek is thought to be from Turkish. Similarly, Aro *ɣăzetă* 'change in coins' and Megl *gazetă* 'counterfeit coins' derive ultimately from Venetian *gaz(z)eta* 'two-cent coin,' via Grk γαζετα 'change' (Banfi 1985: 99; Papahagi 1974: 617) and/or BCMS *gazeta* 'two-cent Venetian coin.'

In some instances, Romance, especially Venetian, lexical items entered the Balkans via Turkish, e.g., BSl *mandža* 'a type of main course made with a sauce of peppers, tomatoes, and/or potatoes' < Trk *manca* 'food, usually for pets' < Vtn prisoner's slang *mangia* 'mucus' < NItl *mangia* 'fodder' (BER III: 645).

As for Italianate/Venetian nautical vocabulary, among the general terms found in the Balkans are Alb *barkë*, Sln, BCMS (esp. Dalmatia), Blg, Trk *barka*, Grk βάρκα, Blg *varka* 'boat, dinghy,' from Vtn/Itl *barca* (see Skok 1971: 113); Alb *rem* 'oar,' from Vtn/Itl *remo*; and Alb *shërok*, Grk σιρόκος 'southeast wind,' from Vtn *sirocco*. More on this terminology, with additional examples, is to be found in §4.4.3, discussed from the point of view of occupational jargon.

### 4.2.1.6 Turkisms and Islam

The role of Turkish in shaping the Balkan lexicon was huge. There are dictionaries of Turkisms in BSCM and Bulgarian that contain 6,878 and 7,427 headwords, respectively (Šklajić 1966; Grannes et al. 2002).[27] While dictionaries of Turkisms

27 We should note that there are also some pre-Ottoman Turkic lexical elements in the Balkans, e.g., ChSl and ORuss *bĕlĕgъ* 'mark, sign, etc.' with variants and reflexes in modern South Slavic and Russian. Such words are variously assigned to Avar, Proto-Bulgar, or Kuman/Pecheneg. As with the non-Greek Paleo-Balkan vocabulary, they attest to ancient contact, and are simply noted as such here.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

for the other Balkan languages (see §4.2.1) are not as voluminous, in some cases owing, perhaps, to the lesser abundance of literary sources or limitations on the sources considered, they nonetheless attest to a profound lexical impact. In fact, it was the presence of this Turkish loan vocabulary that was considered one of the most striking features of the Balkan languages – a key component in Trubetzkoy's 1930 *Kulturwörter*. In the various subsections of §4.3 and §4.4, this influence is treated in more detail, focusing on the ERIC domains and issues of style and register, respectively. Here, a brief overview of some of the non-ERIC vocabulary is both useful and revealing.[28]

All those words that entered the various Balkan languages via Turkish can be considered as Turkisms. Thus, for example, although Turkish *efendi* 'sir' (archaic) is itself from Greek αὐθέντης 'perpetrator' (see footnote 274), its presence in various Balkan languages is counted as a Turkism and not as a Hellenism, since Turkish was clearly the immediate source. The same can be said of Arabic and Persian words that entered via Turkish, e.g., BSl, BCMS, and archaic and dialectal Albanian (*mutatis mutandis*; Hamp 1973) *džiger* 'liver etc.,' which is ultimately from Persian (cf. also *mandža* cited in §4.2.1.4 above). There are also ambiguous cases where it is difficult to determine whether or not a word entered various Balkan languages via a Turkish intermediary, e.g., if Turkish has borrowed from Greek or Romance but the phonology of the item is such that the source of the word in other Balkan languages may be uncertain (cf. Boretzky 1975: 135–169). Thus, for example, Ancient Greek μάνδαλος 'bolt' is the ultimate source of Med/ModGrk μάνταλος, Trk *mandal*, Alb *mandal*, *mandall*, BCMS *màndal*, BSl *mandalo*, etc. The precise route by which this word entered the various modern Balkan languages, however, is moot.

Given that the Ottomans exerted control for periods ranging from one to more than five centuries over all of the Balkan peninsula except some parts of today's Slovenia and Croatia, it is not surprising to find Turkisms in all areas of Balkan vocabulary (see Friedman 2003a and references therein). Thus, just as the influence of Latin was felt strongly on Byzantine Greek as detailed by the list in §4.2.1.2, so too did Turkish influence all these areas in all the Balkan languages. For administrative terminology, Ottoman terms (*mutatis mutandis* with respect to orthographies and regional variants) such as *vilayet* 'province' and *kaymakam* 'governor' are found in all the Balkan languages to refer to Ottoman institutions, and *kaymakam* is still used in Turkish today for approximately the same rank, although *vilayet* is now strictly historical. Terms such as *aga* '[Turkish] lord' (StTrk *ağa*) are likewise found in all the Balkan languages – Grk αγάς, all others *aga* – and remain current, but have a specifically Turkish referent.[29] Similarly, *asker* 'soldier' (Grk ασκέρι,

28 Tietze 1983 provides a bibliographic overview of works on Turkisms in the various languages of the Balkans through the early 1980s. The main focus is vocabulary, but there is also attention to other more grammatical effects as well. See §4.2.1 for references to more recent works.

29 We should note here that many or most Turkisms entered via local dialects and not the Ottoman written language. As a result, differences between the modern StTrk form and the form in various Balkan languages may be due to the Turkish dialectal source. An example of this is intervocalic /g/, an archaism preserved in West Rumelian Turkish but lost or elided in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Rmi *askeri*, Rmn *ascher*) is now archaic or historical and refers to Turkish soldiers.[30] Turkish *barut* remains the word for 'gun powder' in all the Balkan standard languages (Grk μπαρούτι, everywhere else spelled exactly as in Turkish).[31]

In urban commercial life, derivatives of *dukkân*, e.g., Mac *dukjan,* Blg *djukjan*, Alb *dyqan*, still mean 'shop,' while *sokak* 'alley, street' (same spelling in Alb and BSl, BRo *socac*, Rmi *sokako*, ModGrk *sokaki*) means only the more marked, lower, 'alley.' This last shift illustrates Kazazis' 1972 observation that many Turkisms that were not eliminated were pushed down stylistically, as also with BSl *gjol*, from Turkish *göl* 'lake,' for it is now archaic except in Bulgarian in the meaning 'puddle'; cf. also Kazazis 1975 and Sejdiu-Rugova 2017, as well as §4.4.1, on register. Turkisms are still current in many words for everyday objects: *çorap* 'stocking,' giving e.g., Alb *çorap*, Grk τσουράπι, BSl, Jud *čorap*, BRo *ciorap*; *tencere* 'pot; cooker,' giving e.g., Alb *tenxhere*, BSl *tendžere,* Grk τεντζερές, Aro *tengire*, Rmn *tingire*, Jud *tendjere*; in names of foods: BSl, Rmi, Jud *čorba*, BRo *ciorbă*, Grk τσορβάς in features of the physical world, e.g., *hendek* 'ditch,' giving e.g., Blg *hendek*, Mac *endek*, Grk χαντάκι, Aro *endec/hãndac*, Rmn *hindichi/ hendechi/hândechi,* Jud *hendek*.

Since Islam entered the Balkans via Turkish, like Latin and Greek it was also the vehicle of a new religious vocabulary (see §§4.2.1.2, 4.2.1.3, and 4.2.1.4), e.g., *minare* 'minaret,' *cami* 'mosque,' and *imam* '(Muslim) priest' (e.g., Alb *minare*, *xhami*, *imam*, Grk μιναρές, τζαμί, ιμάμης, Mac *minare*, *džamija*, *imam*, Aro *minare, ğimie*, *imam*). Further, some Turkisms associated with Islam, but not specific to it, were adopted into other languages. For instance, BRo *işalá*, Alb *ishalla,* Mac *inšala*, Blg *inšalla* 'hopefully, may God grant it,' from Trk *inşallah* 'if

the Standard and the East Rumelian dialect which contributed to its base. Cf. in this regard BSl *bendisa* but Alb *begendis* from Trk *beğen-* 'like, please.' It is also possible that some of the semantic shifts that occurred in the passage of words from Turkish into various Balkan languages actually took place at the dialectal level in Balkan Turkish either before or during the diffusion. Mollova 1967: 116, cited in Grannes 1988: 248, makes a point for Bulgarian that is also valid for Macedonian, namely that the majority of Turkisms are borrowed without any significant semantic shift. In the case of religious terminology, e.g., *kurban* 'sacrifice' (Arabic *qurban* 'offering, sacrifice,' cf. Biblical Heb *qārbān, qurbān* 'idem') for 'Eucharist,' the adaptation is essentially a cultural one, given the role of kurban in Islam (on the Balkans, see especially Sikimić & Hristov 2007). Moreover, among Arabic-speaking Christians, *kurban* is the term for 'Eucharist,' although Arabic is not a direct source of the Balkan usage. In some cases, a secondary meaning in Turkish became the primary meaning in the Balkans, e.g., Turkish *muhabbet* 'love, affection, friendship, friendly chat' is taken into Macedonian and Albanian as *mu[h]abet* but with only the last meaning. Similarly, Turkish *bahis/bahs* 'topic, subject, investigation, debate, wager' has only the last meaning in BSl as *bas*. Although in general Turkisms are associated with stylistic lowering in the twentieth century (see §4.3.4), in at least some cases a negative expression was made somewhat less so when borrowed (see §4.3.9).

30 In Romani and Aromanian, however, the obsolescence of the term is more recent than in Balkan languages with a longer history of standardization (cf. Friedman 1996c, 2020b, and §4.4.4 below).

31 The Turkish form is ultimately (probably through Arabic) from an earlier Greek form, πυρῖτις, related to πῦρ 'fire,' that with 'stone' understood applied to a mineral that could create sparks in conjunction with steel; however, the modern Balkan forms, including that in Modern Greek, are from Turkish, so that the word is a Turkism in all the modern languages.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

it is God's will,' and Aro *ilealá* 'God forbid' (Papahagi 1974: 675), from Trk *illâ allah*, and Bunis 1999: 629, for instance, notes such phrases in use in Judezmo as *Ala belani versin!* 'May God curse [you]!' from Trk *Allah belânı versin!*, and *Alah shukyur* 'Thank God!' from Trk *Allaha şükür!*, the last element of which is the source of Mac and Rmi *šukjur*, Alb *shyqur*, Aro *shucur*, which all correspond fairly closely to English 'thank goodness!' or 'finally/at last!'.

Turkish even penetrated the realm of Christian religious terminology, which, given the identification of Turkish with Islam, demonstrates that religions as well as languages can show contact-induced change.[32] Thus, as indicated in footnote 29, we find in nineteenth-century Balkan Slavic texts *kurban* 'Eucharist' (Trk *kurban* 'sacrifice'), as well as *kurtulija* 'the Savior' (Trk *kurtul-* 'save'), *sajbija* 'the Lord' (Trk *sahib* 'master'); cf. Gołąb 1960; Jašar-Nasteva 1970; Miovski 1980; Koneski & Jašar-Nasteva 1989; Grannes 1996:9; Graham 2020.[33] Of particular cultural significance is BSl (*h*)*adži*[*ja*], Grk χατζής, Alb *haxhi*, BRo *hagi*[*u*], and Jud *hadji*, all from Turkish *hacı* (WRT *[h]aci*) 'pilgrim' (ultimately from Arabic).[34] While the *hajj* to Mecca is one of the five pillars of Islam, among Balkan Orthodox Christians, and also Ottoman Jews, the title was (and sometimes still is) used for those who make an analogous pilgrimage to Jerusalem. Izmirlieva 2012/2013, 2014 notes that the title was already in use in in the Balkans by the early sixteenth century. We can also note here that the celebration of a Muslim circumcision (Trk *sünet*, BSl and Rmi *sunet*, Alb *synet*), which, in the Balkans, involves a large celebration (as also in Turkey), is often referred to as a 'wedding' (BSl *svadba*, Rmi *bijav*, Alb *dasmë*), referring to the feasting and dancing. Such usage is a calque on Turkish usage, where *düğün* normally refers to wedding feast but can also refer to the feast associated with a circumcision.

#### 4.2.1.7 Great Power Languages and Balkan Vocabulary (Late Eighteenth to Mid Twentieth Centuries)

Beginning with the late eighteenth and early nineteenth centuries, successive and varied waves of lexical influence from non-Balkan European languages entered the Balkans via the dominant languages of the Great Powers.[35] These influences are not directly relevant to the sprachbund since they entered after its formative period and in the context of general Western European and Russian imperial and colonial expansion. Moreover, in keeping with competing European expansive intensions

32 While Gagauz was and is spoken by Orthodox Christians, and the Karamanlides of Anatolia were Turkish-speaking Greek Orthodox Christians, for most of the Balkans, Turkish was a language of both Islam and the state.

33 For more on *kurban*, see footnote 29; though of Arabic origin, in the Balkan context, the word is a Turkism and thus entered the region through an Islamic intermediary (cf. Friedman 2019c). Graham 2020: 191 notes that in early modern religious texts, Turkisms are more common in Christian texts from Bulgaria but not in Franciscan texts from Bosnia. In general, Graham hypothesizes that social contacts between Balkan Slavic and Turkish were more intimate in Bulgaria than in Bosnia.

34 On the issue of Christianisms in Judezmo see Sephiha 1985.

35 For the period under consideration here, these powers were France, Italy, England, and the Austro-Hungarian, German, and Russian Empires.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

aimed at Ottoman territory, the languages of different European powers dominated different regions of the peninsula and at different times.

Thus, for example, in the case of Albanian, Italian influence was stronger in what became the independent state of Albania while German influence was stronger in territories incorporated in Serbia, Montenegro, and (later) Yugoslavia owing to Austrian and German interests (and railroads) in the region and BCMS as an intermediary of innovation. Ideology and timing also played roles. Romanian was heavily influenced – especially in its lexicon and particularly in connection with standardization – by French and Italian in the nineteenth and twentieth centuries (cf. Close 1974), as these languages were both politically prestigious and genealogically related. For similar reasons, Latin, too, was drawn upon. Aromanian and Meglenoromanian were not so influenced, a fact that correlates with a higher degree of core Balkan features in those latter languages than is found in contemporary Romanian, although the causality is connected with social, geographic, political, and ideological factors as well as lack or delay of standardization.[36] Similarly, Russian was a source of lexical (and some grammatical) innovation in Bulgarian during this period. Owing to the fact that Russian itself turned to Church Slavonic for lexical and grammatical enrichment during the eighteenth century, a kind of Russianized South Slavic was reimported into Bulgarian.[37] In the case of Macedonian, owing to the later date and political circumstances of standardization, the influence came from BCMS rather than Russian. With regard to Modern Greek, Ancient Greek, largely via the high-style archaizing Katharevousa variety, has been the source of vocabulary. For Romani, attempts to use Sanskrit or Hindi have met with only limited success, and in parts of the Balkans, Turkisms have been promoted (see Friedman 1989b).

As indicated above, these are more internal issues in the development of particular standard languages, and they are therefore of only minimal concern here (see Friedman 1986b, 2004c on Balkan standardizations). Moreover, since our focus here is on the results of speaker-to-speaker interactions, standardization is less significant for demonstrating the effects of Balkan-internal language, i.e., speaker (see §§3.2 and 3.2.1), contact. An interesting and illustrative example of such differential influence relates to automobile part terminology. For instance, Albanian has *kandelë* 'spark plug' from Italian *candela*, whereas Greek μπουζί, Romanian *bujie*, and Turkish *buji* are from French *bougie*, with both Romance sources meaning 'candle';[38] Macedonian *šoferšajba* 'windshield' is based in part

36 Although Romanian schools were opened in areas where Aromanian was spoken (Brancoff 1905), they do not appear to have had much impact on the spoken language.

37 Thus, for example, the Common Slavic participial system, preserved in OCS, had died out completely in both modern colloquial languages (cf. §§6.2.2.2 and 7.7.2.2.2). A Church Slavonicized participial system became standard in Literary Russian, and parts of this were imported into Bulgarian. Meanwhile, the native descendant of the old present active participle had become a verbal adverb in Macedonia and western Bulgaria, and this form was incorporated into standard Bulgarian, where it is marginal (Leafgren 2019), and standard Macedonian, where it is frequently encountered.

38 Mac *svekjička*, Blg *svešt*, BCMS *svećica* are all native derivatives based on the respective words meaning 'candle,' which all derive from *svet-* 'light.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

on German *Windschutzscheibe*, whereas Romanian and Greek have *parbriz* and παρμπρίζ, respectively, from French *pare-brise*.[39] BCMS, Macedonian, and Bulgarian all use *auspuh* (from German *Auspuff*) for 'muffler,' whereas the other Balkan languages (and Slovene) have native coinages, in some instances translational equivalents, or other borrowings, e.g., Greek σιγαστήρας (lit., 'silencer'), though its source, *silansié*, from French, can also be used, or Albanian (in Albania) *skapamento*, from Italian (versus *auspuh*, ultimately from German but here via Macedonian/BCMS, in Albanian of former Yugoslavia). Albanian *makinë* from Italian *macchina*, versus Romanian *maşină* and BSl *mašina* from French *machine,* all meaning 'car' or 'machine,' is another example.[40] The Romance borrowing *lavazh* '[car] wash' is typical of Albanian in Albania, whereas in Kosovo the native *larje* is typical. Such loans speak more to economic ties in the period in question than to shared culture per se, except insofar as modern technology helps to define shared experiences on the part of members of many speech communties. In terms of the kind of shared culture to be discussed below (§4.3), we can note that *mersi* (from French *merci*; see Popescu 2020 on this word in Romanian, along with a few other Gallicisms) is a colloquial expression for 'thank you' in both Bulgarian and Turkish (as well as Persian and elsewhere). As the language of international diplomacy until World War Two, as well as the vehicle of education in schools sponsored by the *Alliance Israélite Universelle*, French had significant impact on the Balkans, and especially on Judezmo (Şaul 1983). As noted above, for Romanian, both Italian and French were important sources of vocabulary because of the genealogical and therefore ideological relationship. In the nineteenth century, the political prestige of French was especially important (Close 1974). Nonetheless, these phenomena, despite their common source, usually come from independent developments.

#### 4.2.1.8 English Loans and "Internationalisms" in the Late Twentieth/ Early Twenty-First Centuries

Words of wide diffusion that are associated with aspects of current modern culture are often referred to as *internationalisms*, and while the precise definition may be problematic, it is a useful heuristic term.[41] With regard to the Balkans, Friedman 2003a: 30 notes:

> The adoption of so-called internationalisms, i.e., words of Greco-Latinate or West European origin, by the languages of the Balkans has led to a new commonality of vocabulary. This commonality, however, is not one specific to the Balkans but rather reflects a more global West European-based hegemony.

39 The Macedonian is sometimes folk-etymologized as an adjective *šoferšajbna*. Elsewhere, the term is native: BCMS *v[j]etrobran* 'wind defense,' Blg *predno stăklo* and Trk *ön cam* 'front glass,' and Alb *xham mbrojtjes* 'glass of defense.'

40 While the French and Italian are both ultimately from Latin *machina*, in turn from Doric Greek μᾶχανά, as with Turkisms of Arabo-Persian origin, these temporally deeper etymologies are not of direct relevance to Balkan linguistics.

41 For instance, with regard to how many languages are needed to determine "international," or concerning the focus it brings on "national" language groups as opposed to other classifications.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Thus, for example, the lexeme in Alb *batari*, BSl *baterija*, Grk μπαταρία, Rmn *baterie*, Trk *batarya* – all ultimately from English *battery*, cf. Frn *batterie*, Grm *Batterie*, Itl *batteria* – occurs in essentially this form in languages all over the globe, e.g., Hausa *báatìr*, Arabic *baṭāriyyah*, Hindi *beṭri*, Malay *bateri*, etc.[42] In a Balkan context, the fact that many such words have entered from English prompted Friedman's 2011a: 6 observation that "English is the Turkish of the 21st century." His point there, however, is that puristic anxieties about the supposedly pernicious influence of English are misplaced, a point to which we return in §4.4.4.[43] Friedman 2003a: 6 also observes that more people in the Balkans now know English than know a neighboring or co-territorial language, which potentially adds a new dimension to the investigation of the Balkan sprachbund.[44] In this context, it is interesting to note that at the ninth AIESEE (Association Internationale d'Études du Sud-Est Européen) Congress (Tirana, 2004), a large number of papers dealt with such words in the Balkan languages.[45] By the time of the tenth AIESEE Congress (Paris, 2009), however, attention had returned to the investigation of the Balkans in their own context, the importance of which was Friedman's 2011b point (see now, e.g., Niculescu-Gorpin & Vasileanu 2020 on developments with Anglicisms in Romanian). Despite more than a century of borders, Balkan language contact continues, and the Balkan sprachbund is an on-going phenomenon. At this point in time, global terminology adds little to our understanding of the structural and historical dimensions of the Balkans as a linguistic convergence area; its study is more concerned with the respective standard languages of this and other parts of the world, and while these terms have a bearing on issues such as language attitudes and purism, our observations in this section suffice for our purposes here.[46]

42 The non-European forms come from Haspelmath & Tadmor 2009. Some languages have borrowed from other sources or have devised other forms, but the point remains the same. The lexical item *battery* is ultimately from Latin, into French, and then to English, but the specifically electrical sense was apparently coined by Benjamin Franklin, and thus emanated fom English.

43 Such anxieties resemble the debate concerning the term *killer language*, insofar as English is presented by purists – in the Balkans as elsewhere – as an existential threat, usually to a majority nation-state language. While language death is indeed a global problem, including in parts of the Balkans, the languages that are actually under threat are local minority languages, and the so-called killer languages are the locally dominant majority languages rather than English, which, as a lingua franca, does not enter into the kinds of everyday dominance as do local majority languages; exception must be made for countries where it happens to be the local majority nation-state language (Mufwene 2005).

44 Nonetheless, at least for the linguistic minorities that live in every Balkan country, multilingualism in the Balkan languages remains highly relevant and crucial for day-to-day existence and long-term prospects; as Friedman 2011b observes, the Balkan sprachbund is still alive and well, and changes resulting from it continue.

45 See, for instance, Lloshi 2004, Mihail 2004, Shehu 2004.

46 In some of the Balkan countries, as elsewhere in the world, there is popular concern over the encroachment of English. This concern extends not just to words themselves but to the graphic representation of words, so that orthographies are a point of contention for some, as indeed they have been throughout the modern era in the Balkans; see §1.4 and the User's Guide, pp. xxxi ff., on Balkan writing systems and orthographic reform. Although such issues are associated with the lexicon, we do not pursue them further here.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.2.2 Entry of Foreign Affixes

Besides fully lexical material that is borrowed between languages in the Balkans, especially as discussed above in §4.2.1 and below in §4.3, various derivational affixes entered into and became productive in different languages. As noted in §4.1, the focus here is on derivation with lexical content, while the more grammatical sorts of derivation and the incorporation of foreign inflexional affixes are covered in §6 (e.g., §§6.1.4.1 and 6.2.1.1), as part of morphosyntax. We use the term *entry* of foreign derivational affixes in order to be neutral on the question of whether affixes themselves are borrowed or are imported into a language as part of a full lexical form and then extracted out of that form. It is clear that some material enters as part of another item and then becomes productive. Our interest lies simply in identifying foreign derivational material from one Balkan language that has come to be incorporated into another Balkan language. For that reason too, we do not – and cannot – go into all the details of derivation in all the languages but rather concentrate on some key instances where borrowing is involved, resulting in a sampling of trans-Balkan contact in this area of the lexicon.[47]

Various languages can be identified as the sources of derivational material in and around the Balkans. Turkish is the biggest contributor, but other languages also play a role. We discuss here, in turn, Latin, Greek, Slavic, and Turkish. The important and somewhat controversial affix *-ica/-itsa,* which has Slavic roots and can be used to derive feminine nouns, is dealt with in §4.3.8, since it also pertains to diminutive formation.

#### 4.2.2.1 Latin

The Latin lexical influence discussed in §4.2.1.2 allowed also for the entry of various Latin derivational suffixes into the Balkans. In some instances, the spread is not from Latin directly but from Latin into one language and from that one into others. We mention here a few of the more widely represented ones.

Mihăescu 1978: 237–238 documents the widespread occurrence of the agentive/occupational noun suffix *-arius* throughout Balkan Latin of the earliest period, i.e., beginning in the second century BCE, based on inscriptional evidence. It continues in modern Balkan Romance in, for instance, Romanian and Aromanian, e.g., Rmn/Aro *căşar* 'cheesemonger' (Lat *casearius*), Aro *cărbunar* 'coalman, coal-deliverer' (Lat *carbonarius*), Rmn *fierar* / Aro *hirar* 'smith' (Lat *ferrarius*), among many others. Similarly, as Newmark et al. 1982: 164 show, Albanian has this suffix productively as *-ar*, and also an extended form *-tar*, e.g., *lopar* 'cowherd' (cf. *lopë* 'cow'), *qytetar* 'city dweller' (cf. *qytet* 'city'), *kopshtar* 'gardener' (cf. *kopsht* 'garden'), *luftëtar* 'warrior' (cf. *luftë* 'battle'), *këngëtar* 'singer' (cf. *këngë* 'song'). Browning 1983: 38–39, gives -αριος as one of the "new suffixes first

47 Discussing derivational patterns in a given language would fall within the scope of what we have called linguistics of the Balkans, whereas our focus here is a matter for Balkan linguistics, to extend the distinction discussed in Joseph 2001a and §3.2.1.7 (and §7.1).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

appearing in Postclassical Greek" that "extended [Greek] vocabulary" and that, in some instances, "became extremely productive," being used with native Greek stems, e.g., μηχανάριος 'engineer' (cf. Grk μηχανή 'machine; engine'). This suffix continues in Modern Greek in the form -αρης / *-aris* (by regular sound change), e.g., in περβολάρης 'gardener' (cf. περιβόλι 'garden'), ογδοντάρης 'octagenarian' (cf. ογδόντα 'eighty'). The related neuter suffix *-arium*, used in nouns of instrument, is found in Greek in a Hellenized form in nouns such as Middle Greek αλφαβητάριον 'alphabet book' and συναξάριον 'catalogue of saints listed by anniversary.' Some of these words entered Slavic from Greek, e.g., SSl *sinaksar* 'synaxarion' (from the Greek), OCS *dinarь* 'dinar (coin)' (Grk δηνάριον, from Latin *denarium*). The CoSl suffix **-arĭ* is found throughout modern Slavic and is sometimes thought to have entered Slavic from Latin via a Gothic intermediary (Skok 1972: 49), but it is, in any case, especially productive in BSl and most of BCMS, e.g., BSl *ribar* 'fisherman' (cf. OCS *ryba* 'fish'), *ovčar* 'shepherd' (cf. OCS *ovьcь* 'sheep'), SoSl *žen[s]kar* 'womanizer' (cf. *žena* 'woman,' Sln *ženskar*, others *ženkar*), BCMS and Mac *političar* 'politician' (*politika* 'politics,' Sln and Blg *politika* > *politik*), etc. We can also note here a specific extension of Albanian use of *-ar* to SW Macedonian in the use of names of people from a given village, which is typical of Albanian but not Macedonian, e.g., *Nestramár* 'person from Nestram' (Vidoeski 1999a: 112 as cited in Friedman 2018a).

Weigand 1926, in discussing the wide range of suffixes in Balkan languages with the shape [-ul-] but with quite varying functions and origins, notes a few instances involving a Latin source, especially Alb *-ull* from Latin *-ulus*; in §4.3.8, there are examples of the related Latin diminutive suffix *-ulla* in Balkan languages.

### 4.2.2.2 Greek

Greek provides some affixal enrichment of its neighboring languages, but the most extensive involvement here comes in the more grammatically related derivation of various nominal and verbal stems. Thus, for example, the Greek aorist marker -σ- is productively attached to Turkish verbal borrowings, which use the 3sg *DI*-past as the base, e.g., Aro *kurdisire*, Rmn *curdisi*, Alb *kurdis*, Mac *kurdisa*, Blg *kurdisvam*, 'wind up, set up, etc.' all from Turkish *kur-* 'idem' with 3sg *DI*-past *kurdu*, WRT *kurdi*.[48] Cf. Grk μπαϊλντίζω < Trk *bayıldı* (WRT *bayıldi*) 'faint' (see §6.2.2.2). The Greek verbal derivational mopheme *-Vz-*, apparent in the word for 'faint,' is borrowed into many Romani dialects as the loan-verb adaptor, with other dialects using *-Vn-* and/or *-ev-* which are also Greek verbal morphemes (-υ/αν-, -αιν-, cf. Matras 2002: 128). Another Greek affix in Romani is the masculine nominative singular marker *-s*, which Greek inherited from Indo-European, but which was borrowed into many Romani dialects, e.g., the Balkan II group, in loanwords, e.g., native *čhavo* 'boy,' but *dajos* 'uncle (mother's brother)' from Trk *dayı* 'idem.' A further affix of Greek origin occurring widely in Romani is the noun-forming

48 See the User's Guide on Turkish morphophonemic transcription.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

suffix *-ima/-ema/-imos,* etc. which competes with native *-ibe[n]/-ipe[n]* in deriving deverbal nouns (occasionally also deadjectival), e.g., native Skopje Arli *ha-* 'eat,' *habe* 'food,' *šuži* 'pretty,' *šužipe* 'beauty,' *mar-* 'beat,' *maribe* 'beating, fight, war' but Ágios Athanásios (Sechidou 2011: 35) *marima* 'idem,' Kalderaš (Boretzky 1994: 276) *marimos* 'idem.' In some dialects, the nominative uses the native marker and the oblique cases use the Greek affix before the native case inflection, e.g., *xaben* 'food,' dat. *xamaske* (Sechidou 2011: 26).

#### 4.2.2.3 Slavic

The productive suffix *-itsa/-ica* is dealt with in §4.3.8, since the most common meaning is diminutive (Asenova 2002: 62). Here, however, we note that this suffix has other uses as well, e.g., in toponyms (often themselves of Slavic origin), e.g., Alb *Goricë*, Grk *Kastánitsa*, Rmn *Dîmboviţa*. The suffix can also derive feminines from masculines, e.g., Alb *gomar/gomaricë* 'donkey M/F,' Rmn *bucătar/bucătăriţă* 'cook M/F.' This latter usage also occurs in WRT (along with the Slavic suffix *-ka*), e.g., Gostivar Turkish *dayo* 'maternal uncle' → *daytsa* 'aunt [*dayo*'s wife]', *Muzafer* 'Muzafer' → *Muzaferitsa* 'Muzafer's wife,' *yalanci* 'liar' → *yalancitsa* 'female liar,' *arkadaş* 'friend' → *arkadaşka* 'female friend' (Tufan 2007:104, cf. also Jašar-Nasteva 1970). This last phenomenon, while not importing grammatical gender in West Rumelian Turkish (WRT), does import a morphological real-world gender distinction that is otherwise absent from Turkish, which must use lexical 'male' and 'female' (or related items) when disambiguation is required, e.g., *kardeş* 'sibling' ~ *erkek kardeş* 'brother,' *kız kardeş* 'sister' (cf. also §6.1.3.1).

According to Asenova 2002: 63 there are more than twenty suffixes of varying productivity of Slavic origin in Albanian and Balkan Romance, e.g., *-ьkъ*, *-ište*, *-okъ*, *-[j]an*, *-ъka*, as in Alb *çunak* 'little boy,' *baltishtë* 'muddy place,' *malok* 'hillbilly (pejorative),' *Shkodran* 'person from Shkodra,' *çupkë* 'little girl'; Romanian *ciorac* 'little corvid,' *porumbişte* 'cornfield,' *bucureştean* 'person from Bucharest,' *românc*ă 'Romanian woman.'

The Slavic suffix *-nik,* which forms agentive and other types of nouns, has entered productively into Albanian and Balkan Romance, which share the innovation of using the suffix for denominal adjectives (and, in Romanian, adverbs, albeit rarely) as well, e.g., Alb *drithnik* 'granary,' *qullanik* 'corn pone,' but also *fisnik* 'noble, of good family,' *sojnik* 'of good family, pure-blood,' *prishanik* 'crazy, cracked, screwy,' Rmn *făţarnic* 'hypocrite/hypocritical,' *târzielnic* 'lazy[bones], slow[poke],' *puternic* 'strong,' *zilnic* 'daily.' As Gălăbov 1966: 307–312 (cf. Asenova 2002: 63; Croitor 2019) points out, the productivity of these affixes attests to intimate contact that began with the arrival of Slavic speakers in the Balkans.

Another influence of Slavic on Albanian and Balkan Romance which in terms of borrowing is lexical but in terms of the receiving languages affects morphological classes is the tendency to assign loanverbs to a particular conjugational class. In Albanian, Slavic verbs tend to be adapted to the sigmatic conjugation, while in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Aromanian they are added to the fourth conjugation, 1SG *-escu* (Nevaci 2003–2004), and in Romanian to the formerly productive class in *-i* (Nedelcu 2013a: 21), e.g., BSl *čuka/čukne/čuknuva ~ čukva* 'hit, knock, etc.' (3SG IPFV/PFV/derived IPFV [Mac~Blg]) > Alb *çukit*, Aro *cicãnescu* Rmn *ciocăni*.

A few Slavic prefixes have also entered Albanian and Balkan Romance, e.g., Alb *po-* in *pomendore* 'monument' (cf. *mendor* 'mental, thinker'), Rmn *prea-* in *preafrumos* 'exceedingly beautiful' (Asenova 2002: 63). Slavic aspectual prefixes in languages such as Romani and Meglenoromanian are treated in §6.2.2.2.

Finally we can note that the Slavic suffix *-av* '-ish,' has been borrowed into some Greek dialects in Epirus and Greek Macedonia as -αβους, e.g., πρασνούλιαβους 'greenish' < πράσινος 'green' + diminutive -ούλη- + -αβους, a calque + borrowing on Macedonian *zelenikav* 'greenish,' where *-ik-* is interpreted as a diminutive affix (Margariti-Ronga & Papadamou 2019a: 139 and sources cited therein). Citing Rempelis 1953: 251, Margariti-Ronga & Papadamou also give ασπρούλαβος 'slightly white,' κοκκινούλαβος 'slightly red,' μαυρούλαβος 'slightly black,' ξινούλαβος 'slightly sour,' and πικρούλαβος 'slightly bitter' for Konitsa in Epirus.

### 4.2.2.4 Turkish

By far the most important source language for derivational material in the early modern period was Turkish. There are seven suffixes with concrete lexical meanings that spread widely in the Balkans. Four of them became productive to varying degrees in at least some of the languages: occupational *-CI*, abstract *-lIK*, adjectival *-lI*, and locatival *-(h)ane*; the other three, the privative *-sIz*, the personal *-man*, and the agentive *-kâr* show interesting developments even though they are usually limited to stems of Turkish origin. We take each of these up in turn.[49]

Occupational or agentive *-CI* in Turkish is found in a range of meanings such as 'one who does X' or 'one who is associated with X,' and is often translatable as English *-er*, e.g., *yolcu* 'traveler' (cf. *yol* 'road'), *lokantacı* 'restaurant owner' (cf. *lokanta* 'restaurant'), and its meanings in neighboring languages are quite similar. Although the suffix takes high vowel harmony (*i*/*ı*/*ü*/*u*) in standard Turkish, the form in the Balkan languages is always front-unrounded (*i*). This represents the WRT situation, in which high vowels in final position are all neutralized to /i/. In Turkish, the alternation *c ~ ç* is determined by progressive assimilation of voicing, and this can also be reflected in the borrowing languages (as discussed in §5.6 and illustrated in Table 4.3). Some languages (Grk, BSl) also have distinct feminine forms based on inherited material (e.g., BSl *kavgadžija/kavgadžika* 'quarelsome man/woman,' with

49 See Grannes 1996: 250–258 (an abbreviated conference paper version of Grannes 1980) on the copying of Turkish *izafet* constructions into Bulgarian.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.3 *Turkish -CI suffix in the Balkans*

| | |
|---|---|
| Alb | -xhi / -çi |
| Bro | -gi/ -ci [-u] |
| BSl | -džija / -čija |
| Grk | -τζης / -τσης |
| Jud | -ği/ -či |
| Rmi | -dži/ -či |

a Slavic feminine suffix *-ka*, and Grk καβγατζής / καβγατζού 'idem,' with a feminine suffix seen in ModGrk φωνακλού 'loud, coarse woman' or γλωσσού 'gossip girl'). In the case of Balkan Slavic and Greek, there is also a terminal desinence that is added for morphological adaptation.

Besides its wide use in Turkish-derived vocabulary, such as Alb *bojaxhi* 'painter, dyer,' Grk μπογιατζής 'painter,' BSl *bojadžija*, Rmn *boiangiu/boiengiu*, Aro *boiagi,* Jud *boyadji* (cf. Trk *boya* 'paint,' *boyacı* 'painter'), Alb *jabanxhi* 'stranger,' BSl *jabandžija* 'foreigner,' BRo *iabangi[u],* Jud *yabandji* (Trk *yabancı*), this suffix has passed over into more general productivity in all of the languages. This productivity is shown by several details of its use: it combines with native roots, e.g., Balkan Slavic *lov-* 'hunt,' the basis for BSl *lovdžija* 'hunter'; Rmn *drâmbagiu* 'Jew's harp-player';[50] Aro *ghelagi '*inn-keeper' with *ghela ~ njela* 'lamb,' whence 'one who prepares lamb [for guests]' (Polenakovikj 2007: 54); Bugurdži Rmi *asjav* 'mill,' the basis for *asjavdžis* 'miller' (Boretzky 1993), Sepeči Rmi *mindžardžis* 'womanizer' from *mindž* 'vagina,' *xoxamdžis* 'cheater' from *xoxavel* 'cheat,' Sofija Erli Rmi *vurdondžis* 'cart-driver' from *vurdon* 'cart' (ROMLEX);[51] Jud *palavra* 'word' gives *palavradji* 'chatterbox.'[52] It also occurs with recent loans, neologisms, and in phrasal and slang formations, e.g., Alb *partiakçi* 'party hack,' Grk ταξιτζής 'taxicab driver,' Mac *fudbaldžija* 'inept soccer player,' Rmn *duelgiu*

50 Strictly speaking, this is a loan from Ukrainian, but it is a popular Romanian folk instrument.

51 The Bugurdži form *asjav* corresponds to general Romani *asijav*, etc. < Pers *āsyā* 'idem'; Boretzky & Igla 1994: s.v. also cite *asavdžija, asvadžija*, with the Slavic form of the suffix. In the case of Romani, the Turkish suffixes discussed in this section generally occur only with Turkish words. Thus, for example, in Ágios Athanásios (AA) Romani (a conservative South Balkan I dialect, Sechidou 2011: 7, 35), only one native stem makes use of this suffix: *bova* 'oven' *bovadžis* 'baker.' Cf. also Balkan Romani *dromadžija* (< *drom* 'road) 'loafer' (Friedman 2020b). Similarly, in both Sepeči Romani (Sep), a Balkan I dialect spoken by Muslims who were sent from the Thessaloniki region to Izmir and Agía Varvára Romani (AV), a South Vlax dialect spoken by Christians sent from Turkey to Greece, both relocations in connection with the 1923 Treaty of Lausanne, there is only a single native formation with this suffix (Cech & Heinschink 1999; Igla 1996 cited in Sechidou 2011: 35), Sep *xoxamdžis,* AV *xoxamdžio* 'liar' (*xoxav-* 'lie,' *xoxamno* 'false'). We can note, however, that AA Romani (Sechidou) also extends the suffix to Turkish words that do not occur with it in Turkish, e.g., Trk *asker* 'soldier' AA (Sechidou) ~ Rmi *askerdži* 'idem.' Romani overwhelmingly favors native derivational elements in these contexts. This is arguably connected with the kind of grammatical compartmentalization described in Friedman (2013b, 2020b).

52 Judezmo *palavra* also entered Greek (παλάβρα) and changed in meaning to 'babble,' and thence into Aromanian, where *palavragi* also occurs as 'chatterbox.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'someone crazy about dueling'; the Greek acronymic base ΠΑΣΟΚ (Πανελλήνιο Σοσιαλιστικό Κίνημα 'the Panhellenic Socialist Movement') gives ΠΑΣΟΚτσής 'an adherent of PASOK,' Albanian *thashethemëxhi* 'gossip-monger' (< *thashë* 'I said' + *e* 'and' + *them* 'I say'), Mac *drkadžija* 'jack-off' (person < *drka* 'to jerk off'). Even in Judezmo, which was not subject to the standardizing ideologies of most Balkan languages, Turkish suffixes could be used as a form of lowering, e.g., *sedakero* 'charity donor' (Heb *ṣ[e]daqa* 'charity' + Sp *-ero*) but *sedakadji* 'beggar' (Bunis 1999: 81).

The Turkish abstract noun formative *-lIK* is the source of Alb *-llëk*, Aro *-lãke/ -lik*, Blg *-lăk*, Grk *-(ι)λίκι*, Jud *-lik,* Mac *-lak*, and Rmn *-lîc/-lâc*, all with the same function. As with *-CI*, this suffix takes high vowel harmony in standard Turkish. In WRT, however, the front/back opposition is neutralized in favor of the back vowel in final closed syllables if the vowel is unrounded, i.e., *i* > *ı*.[53] The suffix occurs in words of Turkish origin, e.g., Alb *pashallëk,* Aro *pashalãke*, Blg *pašalăk,* Grk πασαλίκι, Jud *pašalik,* Mac *pašalak,* Rmn *paşalâk* 'the quality of being a pasha, the territory ruled by a pasha, the high life.' The formative also combines with native and old non-Turkish stems, e.g., Alb *njerëzillëk*, Jud *benadamlik* 'humaneness' (Alb *njerëz* 'people,' Heb *ben adam* 'son of man'), Grk προεδριλίκι 'presidency' (cf. πρόεδρος 'president'), Mac *lošotilak* 'nastiness' (cf. *loš* 'bad'), Blg *vojniklăk* / Mac *vojniklak* 'military service [colloq.],' Jud. *hanukalik* 'Hanukkah present' (Heb *Ḥanukah* 'the Feast of Lights'), Rmn *varvarlîc* / Aro *varvarlike* 'barbarism,' Megl *sotsluk* 'friendship' (Asenova 2002: 62), and they combine with recent loans, e.g., Alb *avokatllëk,* Rmn *advocatlâc* 'advocacy (ironic; regardless of the actual merits of the case),' Mac *asistentlak* 'assistantship' (ironic), Blg *doktorlăk* 'doctorship' (ironic). Many of these forms have a marked stylistic value, at a lower level than the Turkish sources (see §4.4.1).

The adjectival *-lI* forms adjectives from nouns and has the general meaning, as described by Göksel & Kerslake 2005: 194, of "'possessing', 'characterized by', or 'provided with' whatever is expressed by the stem." It is of more limited productivity in general. In Alb *-li*/*-lli* is restricted mainly to words of Turkish origin, e.g., *borxhli* 'debtor' (cf. *borxh* 'debt,' Trk *borç* 'debt'). The chief exception appears to be words that denote inhabitants of certain towns, e.g., *skraparlli* 'person from Skrapar,' *prishtina-li/lli* 'person from Prishtina,' and even hyperforms such as *shkupjanali* 'person from Skopje,' *dibranali* 'person from Debar' (StAlb *Shkup*/ *shkupjan* 'Skopje/person from Skopje,' *Dibër/dibran* person from Debar). There are also substantives of origin in -λη- in Greek, e.g., Βαρνα-λή-ς '(person) from Varna.' Macedonian *-lija* and Greek -λη- show the hallmarks of productivity, occurring with native roots, e.g., Mac *vošlija* 'lousy' (*voš* 'louse'), Grk μουστακα-λής 'mustachioed' (cf. μουστάκι 'mustache,' ultimately from AGrk μύσταξ 'mustache'), and with non-Turkish loanwords, e.g., Mac *pubertetlija*

53 In dialects where *ü* > *i*, this vowel is also neutralized. In Standard Macedonian and the dialects on which it is based, Turkish high back rounded vowel is treated like schwa and lowered to /a/, while Greek and Judezmo lack schwa and use /i/ (and most of BCMS uses *-luk*). See Johanson 2001 and §3.2.2.9 for additional discussion.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'teenager' (ironic), Grk μπεσαλής 'one who keeps his word' (based on μπέσα, a borrowing from Albanian *besa* 'faith; honor,' itself a loanword in all the Balkan languages in direct contact with Albanian). In Modern Bulgarian, *kuražlija* 'having courage/strength' appears to be the only non-Turkism with the affix (Andrejčin 1975). Turkish-origin words also occur in Grk σεβνταλής 'lustful; passionate' (cf. Trk *sevda* 'love, passion'), Mac *kasmetlija,* Blg *kăsmetlija* 'lucky' (cf. Trk *kısmetli*), Jud *(u)gurli* 'auspicious,' *kokuli/kokulu* 'perfumed,' *Salonikli* 'Thessalonian (note that Trk *Selânikli* often implies 'Sabbatean' (Trk *Dönme*)), Rmn *tabietliu,* Blg *tabietlija* 'persnickety, pedantic' (Trk *tabiat* 'nature, habit'), Aro *hairli,* Mac *airlija*, etc. 'lucky, blessed with good fortune' (Trk *hayır* 'good, auspicious').

The most restricted of these suffixes is *-(h)ane*, which forms nouns of location in Turkish, e.g., *kütüphane* 'library' (cf. *kütüb*, learned plural of *kitab* 'book'). It is not at all productive in most Balkan languages, occurring in words with Turkish sources like *kafehane,* now archaic in Albanian except in the meaning 'dirty, rundown coffee house' (whereas *kafene* is ordinary colloquial for pub or coffee-house) or *mejhane* 'tavern' (Mac *meana*, Blg *mehana*; also, pejoratively, 'noisy smoke-filled saloon'). It seems not to occur in Greek. It is productive, however, in Macedonian, in the form *-ána*, being found in words of Turkish origin, e.g., *kafana* 'pub' (ordinary colloquial), *meana* 'tavern' (archaic), with native words, e.g., *pilana* 'sawmill,' and with recent loanwords, e.g., *energana* 'heating plant.'

Other Turkish suffixes occur on a more limited basis, at least in current usage, but deserve special mention here nonetheless.[54] The privative suffix *-sIz* 'without, -less' occurs only with words of Turkish origin, but is found in all the Balkan languages. Macedonian, for instance, currently has such forms as *arsaz* 'crook,' *teklifsiz* 'unceremoniously,' and *ugursuz* 'no-goodnik' (Trk *hırsız* 'thief,' *teklifsiz* 'without ceremony,' *uğursuz* 'inauspicious; rascal'), and Grannes et al. 2002: 540–541 give nearly fifty forms in Bulgarian with *-siz*/*-suz*/*-săz*, many of which, however, are now obsolete, dialectal, or highly colloquial, such as *hărsăz*(*in*) 'soundrel, useless' (Trk *hırsız*), *kapasăz(in)* 'good-for-nothing' (Trk *kapısız* 'gateless; unemployed'), *kitapsăz(in)* 'illiterate' (Trk *kitapsız* 'without a book'), *hairsăz(in)* 'scoundrel, good-for-nothing' (Trk *hayırsız* 'useless, good-for-nothing'). For contemporary Albanian, Snoj 1994: 474 lists only four words with *-sëz*, all with exact sources in Turkish: *sojsëz* '(person) of poor stock,' *apansëz* 'unexpectedly,' *edepsëz* 'shameless,' and *nursëz* 'sad, lightless' (Trk *soysuz*, *apansız*, *edepsiz*, *nursuz*). Judezmo has *apansiz, edepsiz,* and *soysuz* 'of bad lineage' (as well as *soyli* 'of good lineage' based on Trk *soy* 'lineage'). Greek has this suffix in just in a few words, most notably γρουσούζης / γουρσούζης 'ill-fated; bringing bad luck' (Trk *uğursuz*), where *-sIz* ends up as an isolated piece with no real value that merely adds to the phonological "bulk" of the word (like some of the etymological inflections discussed in §4.2.2.6.2). According to ILB 1957, Romanian has only [*h*]*ursuz* (< Trk *uğursuz* cited above). Given the ideology of excluding Turkisms in all the Balkan standard languages, it is not

54 The diminutive suffix *-CIK* is discussed in §4.3.8.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

unreasonable to assume that the suffix was considerably more widespread prior to the twentieth century (cf. Grannes 1969 regarding Bulgarian).

Further elements to mention here are *-man* and *-kâr*. In fact, *-man* is not even a suffix in the strict sense in Turkish, but it occurs in a number of words mainly for individuals, such as *duşman* 'enemy' (from Persian), *kahraman* 'hero' (from Persian), *Müslüman* 'a Muslim' (from Persian), *peşiman*/*pişman* 'penitent' (from Persian), and *tercüman* 'interpreter' (from Arabic). This individual-marking usage is taken up in some of the Balkan languages in whole-word borrowings from Turkish, e.g., Blg *kahraman* and *terdžuman*, Alb *mysliman* and *pishman*, Rmn *duşman*, Mac *dušman, pišman*, and *Muslimán*. Significantly, it is even extended in Macedonian and Albanian, combining like a suffix with some native bases, e.g., Mac *lažoman* 'liar' (cf. *laga* 'a lie,' *laže* '(he) lies'), *grkoman* 'Hellenizer,' and Alb *gjataman* 'tall, lanky person' (cf. *gjatë* 'long'), *pordhaman* 'person who farts a lot; full of hot air' (cf. *pordhë* 'loud fart,' *pordh-* aorist stem of *pjerdh-* 'fart'). Also, Mac has one word, *utman* 'dullard' that is either a Mac creation based on (dialectal) Albanian *ut* (StAlb *hut*) 'owl; dullard' or borrowed as a whole from dialectal Albanian. There may be some secondary influence in each language from the Greek-based suffix *-mán* '-maniac' that occurs in each in, e.g., Mac and Alb *kleptomán* 'kleptomaniac' and *megalomán* 'megalomaniac,' since these words designate individuals too. The Macedonian use of *-man* with names of nationalities, e.g., *grkoman* 'Hellenizer' (also Alb *grekoman* 'idem'), *srboman* 'Serbianizer,' may also belong here. Thus *-man* appears to have a mixed origin in the Balkans, but there has been some Turkish involvement. (For more details, see Friedman 2003a.)

The Turkish suffix *-kâr*, a borrowing from Persian, is agentive, e.g., *hizmetkâr* 'servant' (*hizmet* 'service,' cf. BSl *izmekjar*). It appears to be productive only in Albanian, which combines it with some native or old loan roots, e.g., *mundqar* 'hard worker' (*mund* 'effort'), *grabitqar* 'predator' (*grabit* 'capture, pillage'), *ziliqar* 'envious person' (*zili* 'envy') (cf. Boretzky 1975: 265–269).

We can also mention here the suffix *-lAmA*, in which the *-lA-* is used to derive verbal stems from nonverbs and the *-mA* forms deverbal nouns, as in Turkish *temiz* 'clean, adj.,' *temizle-* 'clean (verb),' *temizleme* 'cleaning (noun).' In Macedonian (and, *mutatis mutandis*, elsewhere in Balkan Slavic), the noun *zavrzlama* 'tangle, plot, meddling, etc.,' from *zavrze* 'bind, twist, knot,' is still in common usage, albeit strictly colloquial. Cf. also nineteenth-century Macedonian *daskalaisa* 'teach' (< /daskal-la-isa/), *ugursuzlaisa* 'behave badly,' etc. (Markovikj 1996), which combine Turkish *-la-* with Greek aoristic -(ι)σ- (cf. §4.2.2.2). Similarly, *čuvadar* 'guardian' combines *čuva* 'protect, keep' with *-dar*, a Turkism of Persian origin used to form agentive nouns. Unlike *zavrzlama*, however, *čuvadar* is archaic.

### 4.2.2.5 Western European Affixes

Continuing in the contemporary period a trend begun in the twentieth century, Western European languages have come to be more widely known in the Balkans, and as a result, lexical borrowings have occurred, in some lexical fields massively.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

This in turn has led to derivational material associated with some of these borrowings becoming available in the recipient languages and taking on new life in their new linguistic environment. We mention just a few here by way of illustrating this relatively recent trend in the Balkans. It should be noted that some of these could be considered internationalisms (see §4.2.1.8) of an affixal nature.

One English suffix that is itself of relatively recent origin in English that shows up in some Balkan languages is *-gate* referring to a public scandal of some sort, often involving politicians.[55] This suffix has gained currency in English and it has spread to other languages, being found, for instance, in Hungarian (Kontra 1992) and in contemporary Russian, e.g., *Putingejt*, referring to President Vladimir Putin's suppression of anti-establishment journalists and the suspiciously timed murders and poisonings of his political opponents.[56] As far as the Balkans are concerned, Joseph 1992b documents it for Greek in 1987 in the forms Τόμπρα-γκεητ and ΠΑΣΟΚ-γκεητ, referring to a scandal involving the then-head of the national telephone system, a Mr. Τόμπρας, who had been appointed by the then-ruling party ΠΑΣΟΚ (see §4.2.2.4); he also cites an example from the Serbian press, *Agrogejt* for a financial scandal involving the agricultural conglomerate Agrokomerc, and more examples can be found in the Bulgarian press, e.g., from 2012, both *SRS-gejt* and *Tanovgejt*, both referring to a scandal involving bugs ("SRS," i.e., 'special intelligence devices') and a Mr. Tanov, then-Chief Executive of Customs.[57] In both the Greek and the Bulgarian examples, there are multiple labels for the scandal in question, suggesting a degree of productivity for the suffix in a way different from what the mere coining of novel instances reveals. Macedonian shows what may mark the beginnings of the entry of this suffix, since an American situation involving espionage and shopping carts is referred to in an online article[58] as "*količkata-gejt*," translating 'shopping cart' with the definite *količkata* but leaving *gejt* untranslated, presumably because it carried some meaning as such.

55 This suffix arose via resegmentation from use of *Watergate* referring to the now-infamous Watergate scandal of the early 1970s that led to the resignation of then-President Richard Nixon in 1974. An extensive list, with more than two hundred entries and a disclaimer of not being complete, giving *-gate* words attested in US and international usage is at http://en.wikipedia.org/wiki/List_of_scandals_with_%22-gate%22_suffix.

56 See https://nv.ua/opinion/putingejt-o-sekretnom-kanale-svjazi-trampa-s-kremlem-1232643.html and https://www.compromat.ru/page_9378.htm); the suffix also occurs in Ukrainian and many other languages.

57 The fuller context for these examples is: *Van'o Tanov si izjasnil slučaja za SRS-gejt* 'Vanjo Tanov has clarified the case of SRS-gate' (from http://www.dnes.bg/obshtestvo/2012/01/26/vanio-tanov-si-iziasnil-sluchaia-za-srs-geit.149758), and series of news items listed under тановгейт (tanovgejt) at https://offnews.bg/tags/тановгейт/. Consistent with its colloquial nature, there is no conventional spelling, in that the '*-gate'* element can be written with a hyphen, or continuously, as, respectively, in these two cases. An example with a separating space occurred with the form *prokuratura gejt* 'Prosecution gate,' which has disappeared from the web since this research was conducted.

58 This occurred in an article at www.vecer.com.mk/default.asp?ItemID=35B88A792DE168428F5AEB62F62C44B8 at the time this research was conducted, but the link is now dead.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Another modern Western European suffix with some extension through the Balkans is the Italian suffix *-ese* in nouns and adjectives of ethnic or geographic origin, e.g., *Francese* 'French,' ultimately from Latin *-ē(n)sis*. A suffix with roughly this form, though consistently with a voiced [z] suggesting perhaps Venetian mediation, is found in Albanian, e.g., *senegalez* 'Senegalese,' *jordanez* 'Jordanian,' *eskimez* 'Eskimo,' and *nepalez* 'Nepali,' among numerous others, and in Greek, e.g., Κινέζος 'Chinese,' Φιλιπινέζος 'Philippine,' and Σκωτσέζος/Σκοτσέζος 'Scottish,' among many others. Cf. also Mac *Kinez* 'Chinese [person],' as in BCMS but Blg *Kitaec* 'idem,' which is from Russian. Romanian, too, shows *-ez* in abundance, especially for more "exotic" ethnica, e.g., *japonez* 'Japanese,' *chinez* 'Chinese,' *somalez* 'Somali,' and *finlandez* 'Finnish,' inter alia, but deriving it directly from Latin *-ē(n)sis* is difficult, so that some Italian, Venetian, or even French or possibly learnèd Latin influence in its development and spread is likely.[59] It may have entered with specifically Italianate nouns, e.g., *Calabrese* 'Calabrian,' and spread from there in each language. Moreover, there are sufficient numbers of other such forms in each language to suggest a good degree of productivity for the suffix; Snoj 1994, for instance, lists more than thirty for Albanian.

As noted in §4.2.1.8 about internationalisms at the word level, none of these suffixes has any relevance to the sprachbund per se, but rather they speak to modern developments with the standard languages.

### 4.2.2.6 Miscellaneous

Besides the derivational material discussed so far that can be categorized insightfully by source and thus use, there are some borrowed elements that come to serve derivational functions but do not fit into neat categories overall; thus, their treatment may have a more scattered feel but they are no less important and no less interesting.

#### 4.2.2.6.1 Prefixes

We have mostly chronicled here various foreign suffixes with a derivational role in various languages, largely because suffixes are far more common in the languages involved. But there are some prefixes that enter on specific words. A case in point is seen in Greek μπαμπέσης 'an unscrupulous or evily clever man,' a transformation of Albanian *pabesë* 'disloyal, dishonest,' from *pa-* 'without' plus *besë* 'word of honor; trust.'[60] Albanian *pa-* also was cited by Sandfeld 1930: 116 as influencing a particular use of Aromanian *fără* 'without' as essentially a word-forming prefix,

59 The phonological development is not entirely straightforward, for instance regarding the *-z-*, and it does not show expected vowel alternations.

60 Perhaps initial μπ- here results from sandhi with a preceding masculine or feminine singular definite article where the final -ν would voice the initial π- (e.g., τον παμπέση => τομ μπαμπέση, whence the free form μπαμπέση-), or it may just be a simple distant assimilation of π . . . μπ (*p* . . . *b*) > μπ . . . μπ (*b* . . . *b*). Note also derivatives such as μπαμπεσιά 'act of unscrupulousness,' μπαμπέσικος 'characteristic of unscrupulous people,' as well as the base word μπέσα from Alb *besë*, which

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

as opposed to its usual prepositional use, in the phrase *cu fără nimfricoşatu suflit* 'with an intrepid soul' (lit., 'with (a) without fear soul'), that occurs in the Codex Dimonie.[61] Sandfeld *ibid*. further suggests that the Albanian use of the preposition *pa* 'without' as a word-forming prefix may be due to external influence, as he compares the similar dual use of *bez* in Slavic. This is in contrast with Balkan Romani, which borrows Slavic *bez(o)* as a preposition but uses the early borrowed Indo-Iranian *bi-* as a privative prefix. One can speculate as well, though, that Turkish might have played a role, since the Turkish privative adjectival formative *-sIz*, as in *şekersiz* 'without sugar' (lit, 'sugar.without'), is a suffix, and thus forms a single word with what it attaches to while also carrying out an adpositional function; it therefore, like Slavic *bez*, provides a model for the same form being both a word-formative and an adposition, just like *pa*/*fără*.[62] Note also, from Slavic, Albanian *kolo/kollo-* and *po-*, e.g., *kollofruth* 'incantation against measles,' *kollotumbë* 'somersault,' *kollofis* 'gulp down, swallow up,' *polem* 'people, crowd (< *len* 'be born'), *pomendore* 'monument, memorial'; Rmn *ne-, răz-*, *po-, prea-,* e.g., *nesaț* 'insatiable,' *răsfoesc* 'leaf through,' *ponegru* 'very black,' *poneagră* 'evil woman,' *preafrumos* 'very beautiful' (Xhuvani & Çabej 1976: 161, 174; Asenova 2002: 63). We can also cite here the use of Grk κοντο- 'short' as a prefix in Aegean and Pirin Macedonian, Thracian Bulgarian, and Aromanian (Papadamou 2019b, cf. BER III:*kunde*). See also §6.2.2.2 on the borrowing of Slavic perfectivizing prefixes. Romani dialects in Kosovo, e.g., Bugurdži (Boretzky 1993: 83) borrow the (Geg) Albanian gerund marker *tuj* (Standard *duke*), which, unlike the Albanian, which is prefixed to a nonfinite form, in Romani is prefixed to the inflected present, e.g., *tuj dzav, tuj džal* 'while going.1SG, while going.3SG,' etc. Finally, see §4.3.7.2.1 for a Turkish intensive reduplicative prefixing construct that has diffused in the Balkans.

#### 4.2.2.6.2 Etymological Inflection

The entry of whole words from one Balkan language into another has led to the situation in which a borrowed element can contain inflectional material. Such donor language inflection can end up as inflectional material in the borrowing language, even if slightly altered, as documented in §6.1.4.1 and §6.2.1.1. However, in what is perhaps the more typical case, donor language material that happens to be inflectional is simply treated as donor language material with no special status; the fact that a part of the word is etymologically inflectional is of no import to the borrowing speakers. The material can thus end up serving no inflectional function in the borrowing language. In some instances it becomes

allows for μπα- to be segmented off within Greek, even if, as it seems, neither it nor πα- is used in other words. See also §4.2.2.4 regarding μπεσαλής.

61 In this nineteenth-century Aromanian text written in Greek letters, published by Weigand 1897: 192–297, *fără* is reinforcing the negativity of *ni(m)-* in *nimfricoşatu*.

62 The Balkan Slavic dual use of *bez(-)* is inherited, as it is found in Old Church Slavonic and elsewhere in Slavic, e.g., in Russian.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

derivational, and in others just part of the "phonological bulk" making up the word;[63] in this latter case, it is of inflectional interest only from an etymological standpoint. We survey here a few such instances involving nominal suffixes; there are some verbal suffixes that are treated in this way but as they have a grammatical value or show effects that are tied to grammar (see §4.1), they are discussed in §§6.2.1.1 and 6.2.2.2.

The best illustration of such noninflectional incorporation of inflectional material comes in forms borrowed from Turkish, mostly nominal case endings that occur in fixed expressions. For instance, Albanian has borrowed Turkish *hava* 'air' to mean 'open air, weather' and for the meanings 'out in the open' and 'up in the air,' it has both the native *në hava* (lit., 'in air') and the Turkish ablative *havadan*, both of which can also be used with native verbs like *mbet* 'remain'or *qëndron* 'stay' to mean 'hover': *mbet/qëndron në hava/havadan* 'hovers.' So here there is no synchronic recognition of *-dan* as an ablative case maker. The Turkish ablative, in this instance with the shape *-ten* due to Turkish phonological processes, is found also in the BSl adverb *hepten* 'totally,' as well as Blg *dipten* 'from the bottom; completely, fully,' and BSl *birden* 'at once (suddenly),' literally the ablative of *bir* 'one' (Grannes et al. 2002; Jašar-Nasteva 2000). This last Turkism has also given rise to calques in Macedonian, Romani, and Aromanian, although the calquing languages use 'once' rather than 'one': *od ednaš*, *taro jekvar*, *di nã oarã* 'at once,' respectively. In addition, Bulgarian *dokuzda,* a Rhodopian dialect expression for 'angry, gloomy, hypersensitive,' looks like the locative of Turkish *dokuz* 'nine' (Grannes et al. 2002). However, a more plausible explanation is that it is related to the Rhodopian *dokuzdisan*, the past passive participle of *dokuzdisam*, a variant of *dokundisam* 'affect, insult' from Turkish *dokun-* 'touch' and thus with the same semantics as English *touchy.*

Another expression like this is one that occurs in Greek, αναντάμ μπαμπαντάμ (also αναντάμ παπαντάμ), meaning 'in the distant past,' in Albanian, in the form *denbabaden*, meaning 'since ancient times, forever' (sometimes spelled *den baba den* or *dem baba dem*, as in the song entitled *(jam) fisnik dem baba dem* '[I'm] a noble since days of old' popularized by Ylli Baka; cf. also the song *Korba Çeço*), and in Judezmo, in the form *anandan babandan* (Bunis 1999: 629). In each case, this is an alteration of a Turkish *anadan babadan* 'mother.ABL father.ABL.'[64] The forms with *-m* show assimilation of the first final *-n* to the following initial *b-* and then a distant assimilation to give . . . *m* . . . *m*. The Albanian presumably underwent a fore-clipping eliminating *ana-*, and the front harmony could be the result of native generalization (cf. Jud *dunyade* 'world.LOC,' cited below).[65] The Judezmo medial nasality (*-nd-*) probably reflects Greek phonotactics pertaining to the pronunciation of medial voiced stops, though Spanish phonotactics cannot be ruled out. Given that the nouns in this phrase include the Turkish ablative case ending *-DAn*, this is

63 See also §4.2.2.5 on the incorporation of a derivational suffix into the bulk of a word.

64 The form *anne* is preferred in Istanbul (Redhouse 1968).

65 It could be that the expression was borrowed using *anne* (ABL *anneden*) followed by progessive assimilation of vowel harmony. Since the form *ana* was the normal one in the Balkans, however, this route is less likely, given the oral nature of the contact.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

more likely borrowed directly from Turkish as a ready-made phrase than constructed by Greeks or Albanians who might know Turkish. Interestingly, though, available lexicographical sources, e.g., Redhouse 1968, Akalın & Toparlı 2005, Ayverdi & Topaloğlu 2006, TDK 1963–1977, TDK 1963–1982, do not contain this phrase, but there are speakers, generally of an older generation, who recognize it; younger urban speakers, however, seem not to know the expression as such.[66]

A phrase from a Judezmo humor column published in Thessaloniki, cited by Bunis 1999: 91, *Yo se una koza ke yok yok dunyade* 'I know something that's like nothing else in the world' (*dunyade* < *dünyada* 'world.LOC [-DA]'), also involves a case ending.[67] In this example, Turkish laws of vowel harmony are violated, which could reflect the kind of Turkish used in some Balkan Turkish dialects or the speaker's relative competence in Turkish. The example is arguably midway between a codeswitch and an expressive part of the speaker's Judezmo competence, i.e., given the speaker's imperfect knowledge of Turkish.

On the other hand, Romani *akanadan* 'from now on' (Romani *akana* 'now' + Turkish ablative *-dan)* is a clear instance of the borrowing of the Turkish ablative case marker *-DAn.* Nonetheless, this is really a matter of derivation rather than case inflection, since the Turkish suffix has not replaced the native Romani ablative in *-tar* in those dialects where *akanadan* occurs. (See also §4.3.3.2 on the copying of the Turkish ablative together with a postposition.)

Other nominal inflection has entered Balkan languages in a similar way, being reanalyzed as simply a part of a stem. Thus for example the English plural *-s* occurs occasionally in some borrowings into Balkan languages. For instance, Greek has the indeclinable forms κλιπς 'clip' and τανκς 'tank' as singulars (thus, e.g., το τανκς 'the tank') but with the -ς reflecting the English plural *-s*.[68] The status of English *-s* as part of the stem is especially clear in the Bulgarian form *kets* for 'sneaker, tennis shoe' (from the American brand name *Keds*), since it can be overtly pluralized as *kets-ove*, with the usual plural suffix. The same phenomenon is observed in the borrowing of Turkish *-lAr* into the various Balkan languages (see §6.1.4.1).

## 4.3 Adding to the Typology of Loanwords: ERIC Loans

Having covered the historical content loans, however briefly, we can turn our attention to the type of loanword that gives the greatest insight, in our view, into

66 For Turkish, it also does not occur in online resources, but it can be found in the occasional song or newspaper column. Turkish also has the expression *atadan dededen* 'father.ABL grandfather.ABL' in the same meaning. The nuance of both expressions in Turkish is folksy or rural and illustrates a value that urban speakers are less likely to express.

67 *Yok* 'not existent' is also Turkish and occurs in most Balkan languages, but it does not occur in Nehama 1977, who gives *duña* only as an apocope for Trk *dünya güzeli* 'the most beautiful woman in the world.'

68 Greek also has instances where the Eng *-s* functions as a plural marker; Joseph 1992e: 80 documents το τεστ 'the test' (singular) opposed to τα τεστς 'the tests' (plural), for example, and singular forms of 'tank' and 'clip' without the *-s* (-ς) do occur, i.e., τανκ / κλιπ, for which plurals with *-s* (-ς) are attested. See also §6.1.4.2 and footnote 168 therein.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the nature and origin of the Balkan sprachbund. We start by elaborating on the term introduced above in §§4.1 and 4.2, "ERIC loans," as it gives substance to a key distinction in the assessment of lexical borrowings, certainly in the Balkans and arguably more widely too. As noted, this new term is actually an acronym, though we write it simply as ERIC henceforth.[69] It is to be taken as a modifier of "loans" or "loanwords," thus standing for "(loans that are) *Essentially Rooted In Conversation.*" The concept emanates from a particular set of conditions under which certain classes of loanwords crossed linguistic boundaries in the Balkans.

ERIC loans are those loans that depend crucially on speaker-to-speaker interaction of an on-going and sustained kind, the sort of contact that can be characterized as intense and at the same time intimate, as opposed to occasional and casual. As noted already (see, e.g., §3.1), borrowing can occur without any speaker contact, as with loanwords from Latin into English or from Old Church Slavonic into Russian and thence Bulgarian, where the medium is literary, rather than conversational, in nature, or as with the coining of new names for new technological or medical or scientific advances using (neo-)Latin or Greek roots. But even when speaker contact is involved, there can be different degrees of contact; this fact is recognized explicitly in the Thomason & Kaufman 1988: 74–76 "scale of borrowability," where the borrowing of different types of linguistic material is said to correlate with different levels of intensity of contact among speakers. It is our contention, consistent with this scale, that certain types of loanwords, especially those embedded in interpersonal discourse and conversational use and those that go beyond simple exchange of information and/or association with goods and products, correlate with the intense, sustained, and intimate contact, in a bi- or multilingual milieu, that is necessary for the formation of a linguistic area with structural convergences, i.e., a sprachbund, as described in §3.2.2.10 and §3.4. (See Gast & Koptjevskaja-Tamm 2022 for another perspective on degrees of borrowability.)

We thus draw a distinction between loanwords that are concrete and informational and rooted in specific areas of interactions related to material culture (foods, goods, and the like), which can pass among speakers under very casual contact situations, and those that are essentially rooted in conversational interactions and which need considerable direct speaker interaction in order to be transmitted across languages – the "object-oriented" versus "human-oriented" interaction distinction, in the terminology of §3.1. In this way we are amplifying upon our speaker-plus-dialect approach outlined in §3.1 and §3.3, applying it directly to the area of lexis. For this study of Balkan lexis in particular, we follow and build on the work of two key scholars before us (though Friedman 1986c (see below) and Joseph 1994d foreshadow this approach): Kjetil Rå Hauge and Yaron Matras, who emphasize, in different ways, a conversational basis for borrowings

69 We intend that *ERIC* be pronounced as a regular word, as it is spelled, much as the technical term of verbal typology *TAM*, standing for "tense-aspect-mood," is pronounced as [t$^{h}$æm] by some typologists.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

that go beyond terms for concrete material objects. Hauge 2002 draws attention to the borrowing of pragmatic and discourse markers in the Balkans and notes various kinship-related practices and terminology as well as some expressive usages that speak to "a considerable degree of intensity of language contacts" in the region, a notion we amplify below. Matras 1998 offers a cognitive basis, founded in bilingualism and conversational interactions, for the borrowing of elements of grammar that he calls "utterance modifiers," and Matras 2009 with its extensive survey of different types of word-classes (along various dimensions – part of speech and function as well as semantic class), observes, in discussing the borrowing of adverbs, e.g., Romance *certu* 'certainly' into Maltese, or Arabic *belki* 'perhaps' into Turkish (and on into the Balkans, see §4.2.4.2.1 below), that "all of these may have a lexical core in structural terms, but the label 'adverb' applied to them is misleading, since we are dealing with relatively grammaticalised items that operate at the interaction level, not at the level of straightforward naming or labelling, which is the property of content-lexical items." Matras thus recognizes a discourse basis for much of the borrowing of items he surveys at the level of grammar, as do we. As he puts it, regarding the diffusing (for him via "replication") across languages of greetings: "the replication of matter around discourse-level, para-linguistic gestures also satisfies the need to simplify the management-apparatus of conversational interaction within the bilingual repertoire, and to establish uniform or at least compatible modes of reacting and intervening at the interaction-management level."[70] It is such discourse-related interaction-based items that constitute a significant portion of our ERIC loans; this is not surprising, under the reasonable assumption that due to their conversational salience and frequency, discourse phenomena might well show a greater degree of diffusibility (i.e., borrowability).

Still, it is not just that ERIC loans are rooted in discourse, but also, as suggested by Hauge's invocation of "intensity" and as the discussion and examples below make clear, that they reflect certain kinds of interactions, including those of a playful, friendly, bantering nature, with good will among the participants in the conversational exchanges. Consequently, we see them as being "sprachbund-consistent," as well as "sprachbund-conducive," since they represent those lexical elements that most directly reflect the sort of language contact that is consistent with the emergence of a sprachbund and conducive to its emergence: contact on a day-to-day basis, in a multilingual milieu, that is sustained and intense, yet rooted in interactions that are mostly good-natured in intent – that is to say, intimate.[71] Of course, speakers in all contact situations interact verbally, and forms that are typically resistant to borrowing do get borrowed in other than sprachbund-consistent/conducive contexts.[72] However, verbal interaction alone is not the

70 On greetings in the Balkans, see below §4.3.4.2.2 and §4.3.10.1.2.2.

71 Recall our insistence in §3.1 on the importance of speaker interaction in understanding the sprachbund.

72 Ratliff 2011, for instance, discusses the borrowing of interrogatives in various languages of Southeast Asia, without the same sort of ERIC-related conditions; see also Matras 2009: 198–199. Ratliff also

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

issue, but rather the nature, the intensity and the character, of the verbal interaction. Thus, the preponderance of ERIC loans in the Balkans is what we see as particularly striking. All of this means, moreover, that these ERIC loans are not just incidental as far as the sprachbund is concerned, but rather are diagnostic signs of the social circumstances that lead to the structural convergence that most linguists take as pointing to a linguistic area, a sprachbund.

We see ERIC loans as adding to existing typologies of borrowings, though intersecting with some of them. For instance, it provides an overarching rubric for the types of contact influence that Hauge mentioned, drawing together expressives, gestures, kinship practices, and pragmatic markers in the discourse, along with much else. Moreover, the notion of ERIC loans cuts across a taxonomy of borrowing by word-type, as is implicit in the presentation in Matras 2009 where "lexical borrowing" and "grammatical borrowing" are treated in separate chapters,[73] encompassing certain types of lexical items as well as certain grammatical categories.

Among other widely cited and fairly standard loanword typologies in the literature, the influential studies by Einar Haugen deserve particular mention. They offer classifications that focus primarily on the form of the loan; Haugen 1950 distinguished between "importation" and "substitution," based, as Winford 2003: 41–46 puts it in his own survey of loanword types, "on the presence or absence of foreignness markers," while Haugen 1953 drew a distinction between "lexical borrowings" and "creations," that is between what may be referred to (Winford 2003: 41–46) as the "imitation of some aspect of the donor model," in the former case, and forms that are "entirely native [with] no counterpart in the donor language" even if based on some nonnative material, in the latter. As Winford notes, these distinctions recall Betz's 1949 categories of *Lehnwort* (loanword) and *Lehnprägung* (loan coinage). We can also mention here the concepts of additive and substitutive borrowing (cf., e.g., Lusekelo 2017) depending on whether a loanword adds to or replaces native vocabulary. These notions are similar to Desnickaja's 1988 cultural-historical and ethno-historical types respectively (see also Kahl 2014). Here we can note that standardization processes of the nineteenth and twentieth centuries in the Balkans (and elsewhere) actually involved substitutive neologisms to erase the effects of borrowing (cf. the discussion of Turkisms in §4.4.1).

Another important typological schema for loans is that of Bloomfield 1933, based largely on the content of the borrowed word. He distinguished between, on the one hand, "cultural borrowings," those arising via the often-mutual exchange, between speakers of different languages, representing different cultures, of terminology associated with those cultures, and, on the other, "intimate borrowings," those not obviously linked to cultural objects and that seep into a borrower's usage

makes the important point, consistent with what is said here, that "basic" as far as lexical items are concerned does not equate to "stable."

73 Admittedly, this division may be a function of the work's status as a textbook on language contact.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

due to repeated exposure on a regular basis. In that way they are tied to interaction, and it is this sense of "intimate" that we summon up in our ERIC characterization. Yet another typology, offered by Hockett 1958: 403–407, focuses mainly on the motivation for the loan. Hockett contrasts "need-filling borrowings," essentially Bloomfield's cultural type, though the motivation of "needing" a word for a (new) cultural item is at issue,[74] with "prestige borrowings," where the motivation is the "prestige" that the borrowing language speakers accord to material from the donor language.

All of these typologies are useful and the positive attention given especially to Bloomfield's and Hockett's over the years is well deserved, as the distinctions they embody are important and real. Moreover, they are applicable to various borrowing situations in the Balkans. For instance, the borrowing of Greek ecclesiastical vocabulary along with what became Orthodox Christianity into Slavic fits well under the rubric of cultural loans, and thus were additive, and the entry of Turkish words into the various languages was not just a matter of the "need" generated by Turkish administrative terminology during the Ottoman period. And, to a large extent, the rather remarkable degree to which Turkisms occurred in Greek of the Ottoman period, at least into the early twentieth century, in what is now Edirne (Grk Adrianoúpoli),[75] as described by Ronzevalle 1911, 1912, for instance, or in Bulgarian, as catalogued in Grannes et al. 2002, and for Macedonian as in Jašar-Nasteva 2001 and Cvetkovski 2017, can be attributed to Turkish being viewed as a fashionable, and thus prestigious, language in Balkan urban centers at the time (cf. Herbert 1906: 152; similarly, Turkish-style clothing was fashionable among Christians in some Albanian urban centers such as Shkodra and Elbasan, cf. Marubi et al. 2009). Turkish remained an urban prestige language in what is now Kosovo and North Macedonia until well into the second half of the twentieth century, and that attitude has persisted among some old town dwellers into the twenty-first century.[76]

Still, these classifications are not without problems. For instance, by focusing on form, Haugen's does not build in the social context for the loans, even though, except in the case of learnèd borrowings (and the like), borrowing implies interaction between/among speakers. Moreover, the types listed above are not

74 Thomason 2001, for instance, makes the important point that no language ever "needs" to borrow a word, as there are always language-internal derivational means available to speakers (including compounding and metaphorical extension, among other processes), even if the option of borrowing a word along with a new object is an easy and thus often attractive solution to the problem of naming said object.

75 The Anglicized form *Adrianople* was in common use until the Turkish Postal Service Law of 1930. See now Joseph 2019c on the Greek of this city during the Ottoman period, based on Ronzevalle 1911, 1912.

76 The identity of Edirne Greeks extended to adopting Turkish attitudes toward dress. For instance, regarding *šapka* 'hat,' Ronzevalle 1911: 318 notes that "*N'est employé par les Grecs qu'ironiquement pour imiter le dédain des Turcs pour cette coiffure des Européens, car les Grecs ont le mot καπέλου*" 'it is used by Greeks only ironically to imitate the disdain of the Turks for this headcovering of the Europeans, because the Greeks have the word καπέλου.' We can note here that *šapka* itself is a Slavonicism, and that at the time loyal Ottoman subjects all wore the fez.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

necessarily discrete – a cultural/need loan might be undertaken for reasons of (Hockettian) prestige or be associated with (Bloomfieldian) intimate contact, as seen in the shifts in religious vocabulary discussed in §4.2.1.6. Furthermore, non-cultural/nonneed loans do not always involve prestige, at least not obviously so; for instance, as described in Joseph 1985b, drawing on Fourikis 1918, the Greek of Megara (in the area of Corinth) shows the borrowing of the Albanian diminutive *–zə* in forms such as λιγάζα 'a little' (cf. Greek λίγο 'little (N); a little'), even though it is not at all clear that speakers of Albanian or their language ever had any prestige in that, or indeed any, part of Greece.[77] Moreover, as this Albanian loan in Greek shows, along with the numerous other loans discussed below, this sort of borrowing at an intimate level can go in all directions, contrary to Bloomfield's view; Bloomfield 1933: 461 saw intimate borrowing as essentially a one-way process: "intimate borrowing is one-sided … the borrowing goes predominantly from the upper language to the lower." Recognizing conversational interaction as the basis for the loans, instead of local prestige and relative relations of "upper" and "lower," makes the two-way, bi-directional nature of these loans seen in the Balkans readily understandable.

What these typologies are missing (as Matras 2009 recognizes) is the full dynamics of the environment in which borrowing occurs and the medium through which borrowing takes place. This is a particular concern for the Balkans, since the lexical side of the Balkan sprachbund is only one dimension of the contact-related effects, inasmuch as there is massive structural convergence evident too.

What is needed, therefore, is the recognition of a type of loan phenomenon which is consistent with what is known about contact in the Balkans, the contact that gave rise to the structural convergence that in part defines the sprachbund, namely sustained, intense, intimate contact among speakers, with multilateral, multigenerational, mutual, multilingualism. Identifying a class of ERIC loans does just that, as they are based on the mutual interaction, specifically on conversational interaction, between speakers. Thus, we see our "ERIC loans" as extending existing typologies, especially as to the notion of "intimate loans," but without entirely endorsing the traditional and still quite prevalent rubrics for analyzing loans and all that they entail, such as Bloomfieldian "one-sidedness."

These ERIC loans are similar in certain ways to Trubetzkoy's culture words, a class of loanwords that reflect a shared cultural milieu for the languages in question. In a sense, ERIC loans are a type of culture words, namely those associated with the culture of conversational interaction among speakers. Their connection to Trubetzkoy's conceptualization of the relevance of loanwords to recognizing a sprachbund, then, provides a basis for their characterization here of being both "sprachbund-consistent" and "sprachbund-conducive."

As an extension of this comment on the relevance of loanword evidence, culture words and ERIC loans explain why the label of "Balkan language" is not

77 See §4.3.8 for more on the borrowing of diminutives in general and of Albanian *-zë* in particular. See also Matras 2002: 249–250 on covert prestige.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

a typological notion as far as the sprachbund is concerned. That is, while it has been claimed that English could be counted as a Balkan language (see Aronson 2007), since it shows case mergers, a volitionally based future, retreat of an infinitive (though with the emergence of a new one with *to*), and so on,[78] such a claim only makes sense if understood as a typological statement that the language has such and such features. It is meaningless historically (as Aronson himself of course recognizes), and in examining the Balkan sprachbund, we are examining the history of the languages in the area, how the convergences emerged, what conditions gave rise to them, and so on. Trubetzkoy's culture words and our ERIC loans are part of the determination of the sprachbund, and a language like English can be excluded not based on the features it shows, but on the fact that it shows no signs of participation in the diffusion of culture words as well as ERIC loans. Similarly, arguments that the Balkan sprachbund is just part of a larger European convergence zone (e.g., Haspelmath 1998; see §3.4.1.3 and §7.7.2.1.5) fail on this lexical dimension. In this sense, then, the lexicon can prove diagnostic as to sprachbund "membership."

Moreover, as discussed further in §6.2.5.11 and Chapter 8, the conversational nature of ERIC loans serves as a basis for understanding various Balkanisms that are more pragmatic in nature, as with inferences about information source ("evidentiality" – see §6.2.5). Such effects depend to a large extent on conversational interaction between speakers and the inferences that are drawn in the full context of conversation.

To return to ERIC loans per se, the acronym "ERIC" is motivated in two ways. First, as alert readers may have already noticed, it serves as a suitable homage to Eric P. Hamp, who not only provided the authors with invaluable guidance, insight, and advice on the Balkans countless times in past decades, but also offered the same level of sagacity and wisdom to untold numbers of Balkan scholars over the years through his literally hundreds of articles and presentations treating all aspects of Balkan linguistics. We are pleased to be able to offer this homage since our mentor Eric passed away in 2019, on February 17.

However, this is not just an idle way of honoring a scholar and mentor we learned much from. There is a second motivation, namely, that as the brief review of loanword typology above demonstrates, there is a need for distinguishing between loans that take place under sprachbund-conducive conditions and those that take place under casual contact situations. In our view, face-to-face interaction, of the sort that would necessarily have occurred under the intense and on-going contact among speakers in the Balkans, when coupled with multigenerational, multilateral, mutual multilingualism, is essential for creating and propagating the structural

78 The same could possibly be said about Spanish (though not Judezmo) since Spanish has clitic doubling, case mergers (ultimately, total case loss), marking of a personal direct object, and so on, all features found in the Balkan sprachbund. Cf. also Vaux 2002 on putative similarities between the Balkan sprachbund and its relationship to Armenian and (Anatolian) Turkish. Vaux's numerological approach, use of dubious sources, and occasional achronicity all render his arguments deeply problematic, at best. For a serious and nuanced treatment of Anatolia as a linguistic area see Donabedian & Sitaridou 2021 and the sources cited therein. Cf. also Stilo's 2018 Araxes-Iran sprachbund, which includes Armenia and eastern Anatolia.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

convergences typically taken as diagnostic of a sprachbund.[79] The fact that certain kinds of loanwords occur in such a social milieu is an additional factor, and it means that such loanwords can be both another indicator of contact conducive to the formation of a sprachbund and, therefore, a result of such contact. Thus, these are loans that tell us about speaker contact and about the sociolinguistics and sociohistory of the region. For that reason, we give our primary emphasis here to these borrowings and in the sections that follow we survey different types of ERIC loans in the Balkans.

Although extensive documentation of these loans comes in the following sections, a brief, and particularly telling, example of some ERIC loans in the Balkans, can be offered here: the entry of Turkish words into Macedonian, where, as described in Friedman 1986c, virtually all categories of lexical items, covering virtually all sectors of the vocabulary, have been affected:[80]

> The large number of Turkish lexical borrowings belong to all levels of vocabulary and almost all parts of speech, e.g., *džeb* 'n. pocket' (*ceb*), *bendisa* 'v. please' (*beğen-*), *taze* 'adj. fresh' (*taze*), *badijala* 'adv. for nothing' (*bâdihava*), *ama* 'conj. but' (*amma*), *karši* 'prep. opposite,' (*karşi*), *ič* 'pron. nothing' (*hiç*), *sikter* 'excl./interj. scram' (*siktir*), *kešk*i 'part. if only' (*keşke*). The only Macedonian traditional part of speech lacking Turkisms is the numeral, although there are Turkisms in numerical expressions, e.g., *čerek* (*çeyrek*) 'quarter,' and Turkish numerals in other parts of speech, e.g., *bešlik* (*beşlik*) 'five-grosch silver coin' . . . Turkish vocabulary has penetrated every facet of Macedonian life: urban and rural, e.g., *duḱan*, 'shop' (*dükkân*), *sokak* 'street, alley' (*sokak*), *ambar* 'barn' (*hambar*), *endek* 'ditch, furrow' (*hendek*); man-made and natural, e.g., *tavan* 'ceiling' (*tavan*), *šiše* 'bottle' (*şişe*), *zumbul* 'hyacinth' (*zümbül*), *taftabita* 'bedbug' (*tahtabiti*); intimate and abstract, e.g., *džiger* 'liver, lungs' (*ciğer*), *badžanak* 'brother-in-law (wife's sister's husband)' (*bacanak*), *rezil* 'disgrace' (*rezil*), *muabet* 'conversation' (*muhabbet*).

The ERIC loans to be surveyed here fall into these sorts of categories, and more. We classify the subtypes of ERIC loans in Table 4.4.

It can be noted that some of these ERIC types are not lexical borrowing in the strict sense. We include the spread of expressive phonology here and also onomatopoeia because, for one thing, these phenomena spread under the intimate and intense contact we see as crucial to the Balkan sprachbund and, for another, they have lexical effects, adding to or altering the shape of a given word as listed in the lexicon. Moreover, they reflect the fact that what goes on in interpersonal discourse, and conversational interactions in particular, is not just for the exchange of information; there is an emotive and expressive side as well. Further, some phenomena of a morphological nature, in particular vocatives (§4.3.5, §6.1.1.4),

79 This is in keeping with the views of Thomason & Kaufman 1988 on the contact conditions necessary for a sprachbund to arise; see §3.2.2.10.

80 A similarly wide range of lexical classes can be observed for Turkish loans in all the Balkan languages, especially during the nineteenth century. A similar point is made by Sonnenhauser 2015. For Bulgarian, see also Grannes 1988, and for the other relevant Balkan languages, see the sources cited in §4.2.1.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.4 *ERIC loanword categories*

- Kinship terms
- Numerals
- Pronouns
- Adpositions
- Pronouns
- Negatives
- Complementizers
- Discourse elements (connectives, attitudinal expressives, interjections, gestures)
- Vocatives
- Onomatopoeia
- Reduplication (especially of an expressive nature)
- Expressive phonology
- Diminutives
- Taboo expressions
- Idioms (and phraseology more generally, even shared proverbs; also, isosemy, discussed below)
- "Intimate cultural items"[81]
- Secret languages, trade languages, and jargonistic usage

and even a morphosyntactic and pragmatic nature, in particular ethical datives (§6.1.1.2.5), evidentiality (§6.2.6), and narrative imperatives (§7.8.2.2.8), fit into the conversationally based rubric that ERIC loans determine, vocatives through their connection to address and thus conversation, ethical datives through the speaker and interlocutor involvement that they mark, evidentiality through the importance that speakers accord to information source or speaker attitude in conversational exchanges, and narrative imperatives through the vividness and immediacy they impart to oral-based narratives.

We include as well some processes that apply to and operate on words, such as reduplication, given that they augment both the form and the meaning of particular lexical items. Interestingly, and importantly, as others have done,[82] Hauge 2002 blends the expressive with the reduplicative in including expressive reduplication with *m*-, discussed below in §4.3.7.2.2, as relevant to his treatment of Balkan "pragmatic and paralinguistic isomorphisms." We note, though, that he does so without a framework that connects the pragmatic with the paralinguistic, except via "intensity of contact," a connection that our concept of ERIC loans does achieve.

Further, in the case of idioms and phraseology, the traditional notion of calquing is evident, inasmuch as these contact-induced cases involve native-language

81 This is a notion that refers to such phenomena as the use of Muslim terms in nineteenth-century Macedonia for Christian concepts as noted in §4.2.1.6.

82 Cf. also Joseph 1984b and, especially, Joseph 1994b, where *m*-reduplication is brought in under a more general consideration of lexical borrowing in the Balkans.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

material arranged according to patterns in another language as the models. As discussed in §3.2.1.7, the distinction between traditional borrowing and traditional calquing is not an issue for our focus here, in that in either case, foreign language material, whether the form itself or just the semantic framework for a form, finds its way into the recipient language. We thus discuss in §4.3.10 several parallels that fall under the rubric of *isosemy*, a term for parallel semantic structuring in a contact situation.[83] The one way in which the distinction does matter, however, is that in calquing (again, see §3.2.1.7), one has to assume at least minimal knowledge of the donor language and its structure and/or morphemic divisions on the part of a recipient language speaker. Since multilingualism is a crucial condition for the emergence of a sprachbund, the fact that calquing points to bilingualism is a crucial indicator of sprachbund-conducive/consistent conditions.

Returning to ERIC loans, we observe that many of them are members of closed lexical classes, including function words, and many represent vocabulary domains generally held to be somewhat resistant to borrowing. Swadesh 1950, for instance, developed this notion and with it, a list of 207 words for what he saw as generic and pan-cultural concepts that seemed particularly likely not to be borrowed.[84] Many have questioned the underlying assumption behind Swadesh's list, especially, Matras 2009: 166–167, who says "given that the list is rather short, it is difficult to use it as a basis for statements about the stability of either individual semantic domains or word-classes." Nonetheless, it is much cited as an important early statement about borrowability and offers Swadesh's educated opinion on the resistance of certain words to replacement by borrowing. Moreover, Thomason & Kaufman 1988, in their "borrowing scale," assign to level three contact (out of five levels) – the level of "more intense contact"[85] – various of the word classes listed above, including "adpositions . . . personal and demonstrative pronouns and low numerals," as well as function words more generally, a class which would take in complementizers and negation markers. And, as published in 2009–2010, work conducted at the Max Planck Institute for Evolutionary Anthropology, through the development of a Loanword Typology database,[86]

83 Feuillet 2012 uses the term *iso-sémie*, and Kyriazis 2012a uses *isosemy*, adopted here. Sandfeld 1930: 88 writes of words that offer "*le même sémantisme*" ('the same semantics'). Weinreich 1968: 48, footnote 18 notes that this phenomenon is referred to by Haugen 1950: 219, as a *semantic loan* or *semantic extension*, as a type of *loan-shift*, and by Betz 1936, 1949 as a *Lehnbedeutung* ('loan meaning').

84 Swadesh's interest, as part of his "glottochronology" methodology (cf. especially Swadesh 1950, 1955), was in words most likely to be inheritances, which could be used to judge the extent of lexical replacement over time, and thus offer a measure of the time-depth for the divergence of related languages. His list, though, has come to be used for other purposes. (See also Grant 2001.)

85 Their levels are characterized by "casual contact" (level one), "slightly more intense contact" (level two), "more intense contact" (level three), "strong cultural pressure" (level four), and "very strong cultural pressure" (level five), characterizations which the authors recognize are somewhat vague, even if they have common sense interpretations. The notion of *pressure*, however, potentially implies a hierarchy different from the mutuality that we posit.

86 See, for instance, Haspelmath & Tadmor 2009a, Haspelmath & Tadmor 2009b, Tadmor 2009, Tadmor et al. 2010, and related work available on the web that discusses the Leipzig Loanword Typology Project (https://wold.clld.org/).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

offers a statistical basis for borrowability, and, significantly for the concept of ERIC loans, some of the categories listed under our rubric fall within the semantic spheres claimed as least likely to be involved in borrowings. Finally, while Matras 2009: 193ff. surveys cases of borrowing involving word-classes believed to be generally resistant, he is comfortable with the idea that such classes resist borrowing just under "usual circumstances" (our term, not his), since he says (p. 165) that "often . . . the counter-example . . . can be explained as resulting from a local, language-particular constraint that impedes the realization of common patterns in a particular instance." Thus, the occurrence of such loans in languages in the Balkans speaks to the nature, and intensity, of the local contact situation.[87]

Of course, not all closed class items are resistant to borrowing, as certain discourse-related reasons can favor the borrowing of some such words. In all instances, though, the dynamics of discourse, and of conversational interaction, are key to understanding the borrowing that occurs.

In what follows, we survey the classes of ERIC loans identified above in Table 4.4. We start with the closed class items, both lexical and grammatical, and then move towards the more expressive end of the ERIC typology including shared formulaic usage and phraseology. Some of these subtypes receive fuller treatment elsewhere; vocatives are discussed further in §6.1.1.4, and expressive phonology in §5.7. Still, they are included here as they contribute an important dimension to the overall concept of ERIC loans and thereby to the sprachbund. The forms given here, then, are intended to be more illustrative of the patterns of borrowing than exhaustive; still, we aim to provide as full a range of the languages showing these borrowings as possible based on available sources. Their existence is their relevance for the sprachbund, but so too are the sheer numbers of such loans in the region; Matras 2009 gives examples of loans here and there from around the world that are like some of the subtypes of ERIC loans surveyed here, but what makes the Balkan situation so striking is that so many different types of conversationally based loans are represented in the region in such a concentrated way – in our view, that strengthens the claim that they represent a dimension that was crucial to the formation of the Balkan sprachbund.

87 We do not consider it necessary to comment on all of the claims in Haspelmath & Tadmor 2009a and others (see preceding footnote 86), and so just note, where appropriate, that certain categories exemplified here run counter to their expectations. Since not all of their claims fit into our categories easily, we note here that the Balkans offer evidence bearing on their claims about words for "emotions [and] values." Such words, they say, are relatively resistant: they are borrowed in only 24 percent of the languages in their sample, and thus rank eighteenth out of twenty-four categories included. However, Turkish words pertaining to emotions and values abound in Balkan borrowings: *inat* 'stubbornness' occurs (or occurred) in all of the Balkan languages, *rezil* 'disgrace' was likewise widespread; the culturally loaded term *besë* in Albanian, meaning 'trust, loyalty, faith, oath' – a concept that is a cornerstone of Albanian clan-based society – has been borrowed into all the Balkan languages (although it may now be an archaism in some of them); the Turkish *sevda* (of Arabic origin) 'love' also occurs (or occurred) in all the Balkan languages and forms of the Turkish Arabism *keyf* 'pleasure, delight, joy, merriment' also occur in all the languages: Alb *qejf,* Aro *kefe,* Blg *kef*, Grk κέφι, Jud *kyef*, Mac, *kjef*, Rmn *chef,* as well as BCMS *ćef*, as do a number of other such items.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

A cautionary note is needed. When convenient, we often cite modern standard forms from the languages involved; especially in the case of Turkish, as Johanson 2002: 108 cogently points out, the relevant source was often colloquial Balkan Turkish, in which words, meanings, and sounds can differ greatly from modern standard forms. We recognize this problem but argue that tracking down every pre-modern and dialectal form is beyond the scope of what we are trying to demonstrate here. Thus any serious etymological work that might emerge out of our discussion below needs to consult specialist literature regarding the details on the likely earlier source words.

### 4.3.1 Kinship Terms – General Concerns, Exemplified with '(Grand)Father'

Kinship terms are universally recognized as a closed set of lexical items bound to their cultural context. As such, they tend to be resistant to borrowing, and to be sure, the terms 'father,' 'mother,' 'child,' 'wife,' and 'husband,' occur on the "Swadesh list" (see footnote 84).[88] Such terms would be covered by level 3 ("more intense contact") in the borrowing scale of Thomason & Kaufman 1988: 74–76. Matras 2009: 169 argues that immediate kin terms resist borrowing, since they are part of the "general stability of concepts pertaining to the immediate surroundings [which includes] orientation in space, time and quantity, the private domain of mental and physical activity, and the nearest human environment (body and close kin)"; he refers to this characterization as a "proximity constraint" and gives a hierarchy (p. 161) in which "more remote kin > [=are more easily borrowed than] close kin." The sort of contact needed for the acceptance of borrowed kinship terms into wide usage would thus be intense and sprachbund-conducive and thus associated with ERIC loans.

And indeed, in the Balkans the borrowing of kinship words has taken place numerous times, involving several different pairs of languages and a variety of terms, both close and more distant kin. In many instances, where it is possible to tell from an etymological standpoint, the source of Balkan kin-term borrowings is Turkish, but almost all of the languages – Slavic, Greek, and Albanian in particular – figure in kin-term borrowing. The need for etymological caution here is dictated by the fact that the presence of similar-looking kin-terms in some of these languages may well represent nursery terms that were independently arrived at in each language. For example, while Aromanian *tatã* (Cuvata 2009: s.v.) looks like the widespread Greek τάτας and Epiros Greek τάττος (both cited, with references, by Papahagi 1974 and Vrabie 2000) and the Albanian *tatë* 'papa' (given as "colloquial" in Newmark 1998) and Macedonian *tatko*, voc *tate* (although Bulgarian has *bašta*), the cross-linguistic prevalence of *CaCa* words, especially with coronal (and labial) consonants, for intimate kin-terms (so-called "nursery

88 Cf. the use of Latin *mater* and *pater* in affected English, however, which illustrate the manner in which literary sources can influence spoken language. The importance of the social construction of kinship has been elucidated by Schneider 1984, among others. See Mitterauer 2000 for a discussion of the medieval origins of modern Balkan kinship terminology.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

words," and cf. English *dada*, *daddy*, *dad*) means that any given instance of form like this in the Balkans could simply have arisen within that language on its own; contact and borrowing need not be involved.

Still, there are cases where borrowing must be the explanation for a particular kin term in a given language. For instance, Albanian, Aromanian, Greek, and Romani all have *babá* (Grk μπαμπάς; also Geg *bábë*, with typical retracted stress in Turkisms, and Bugurdži *bábi* (Boretzky 1993: 34), probably influenced by the Geg definite) for 'father'; even though seemingly derivable independently in each language as a nursery word, it most likely is a borrowing from Trk *babá*, since the stress placement, in Greek at least, points to Turkish as the source, and in Albanian, this word uses a Turkish plural marker: *baballarë*.[89] Moreover, contemporary Macedonian sources give *baba* 'father' as an archaism (Velkovska 2003: s.v., Derebaj & Filipov 2019: s.v.), suggesting that it was in wider use in earlier times, and similarly, Bulgarian lexical resources cite it as dialectal (see Grannes 1996: 164).[90]

It should be noted too that in each of these cases, the languages do have other words in use, e.g., for 'father' Greek has πατέρας, Aromanian has *néni*, and, rarely (mostly in speech communities in Greece) *patéra*, Macedonian has *tatko*, and Albanian has *atë* as well as a form *lalë*, given by Newmark 1998: 437 as having one meaning, marked as "colloquial," that is relevant here, 'young father: daddy' (also in Çabej 2014: s.v.); interestingly, this last word has many other kinship-related meanings (see §§4.3.1.1, 4.3.1.2, and 4.3.1.4).[91] At least one of these is a clear loanword, namely Aro *patéra*, from Greek, and Alb *lalë* is claimed to derive from Trk *lâla* 'manservant assigned to the care of a child' (ultimately from Persian). The remaining forms other than the Greek may well have nursery-like origins if not as a new creation within the language itself then in an earlier stage; Aro *neni*, for instance, derives from a presumed Lat *ninna*, which originated as a nursery form.

Importantly, by way of showing another dimension to the examination of these words, it can be noted that these various forms are (or at least at one time were) stylistically differentiated from the loanwords, much as nursery *daddy* is from formal *father* in English, with the loanword being relegated to a more intimate and colloquial register of use.[92] While Albanian *atë*, for instance, can simply be

89 See §5.6 on the phonology of the Turkish *-lar* suffix in Albanian, and §6.1.4.1 on the distribution of Turkish plurals in the Balkans.

90 In some of Balkan Slavic the loanword *baba* 'father' coexists with inherited *baba* 'old woman, grandmother.' Note too that there are examples, e.g., in Macedonian folk poetry, where the vocative, in *-o*, has impinged on the form used as a subject so that it too ends in *-o*. BSl also has *babajka* 'father,' a contamination from *majka* 'mother' (BER I:s.v.*baba*). We note too that Vasmer 1950–1958: s.v. gives Russ *babá* and *babáj* as dialectal loans from Turkic.

91 The Albanian and Macedonian forms are themselves likely to be derived from nursery words, though the Albanian especially has fairly direct cognates in Ancient Greek ἄττα 'daddy,' Gothic *atta*, Hittite *attas* 'father,' and Slavic *ot*(*ьсь*) 'father,' which together suggest a Proto-Indo-European **at-* 'father' (but plausibly a nursery term for 'daddy' (Pokorny 1959: 71; BER VI: *otec*)).

92 See below, §4.4 for more on stylistic differentiation. The ModGrk μπαμπά- vs. πατέρα- distinction is a stylistic one based on context of use and degree of intimacy, and it is reminiscent of a different but nonetheless somewhat similar contextually based distinction in Homeric Greek between the vocative ἄττα, used as an honorific address to elders, and the vocative πάτερ 'father' (literally; see

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'father,' it is also found in metaphorical use in *atdhe* 'fatherland.' And, the current Macedonian and Bulgarian situation reveals the nature of lexical competition in words for 'father': for 'father' there is *otec*, which is used just as 'father' in Church titles (OCS had the broader meaning; see BER IV: s.v. for discussion). In Macedonian, the vocative form *tate* is the intimate term akin to *daddy*, while *tatko* is the "normal" (unmarked) register word for 'father,' and is the basis for the derivative *tatkovina* 'fatherland.' Derivatives of the *tat-* root are widespread in Bulgarian (BER VII: s.v.), including *tatkovina* for 'fatherland,' but 'father' per se is *bašta*, which, etymologically, is related to 'older brother' (BER I: s.v., cf. the next paragraph).

The terms for 'father' in the older generation, that is, 'grandfather,' also show some contact effects. Among the meanings for Albanian *lalë/lalo* is 'grandfather,' according to Meyer 1891: 236; but Mann 1948: s.v. gives only 'elder brother, (paternal uncle) daddy, godfather, term of endearment for bishop.'[93] Çabej 2014: s.v. glosses *lalë* as 'older brother, father, brother-in-law' and notes the comparisons with Greek, Aromanian, Turkish, and BCMS. He notes those scholars who claim it as a Turkism in Albanian and those who consider it a nursery word. He also observes that it is more common in the south and center of Albania than in the north and is used as a nickname for inhabitants of Myzeqe (cf. the same for inhabitants of Vojvodina in BCMS, usually with an exaggerated rising tone). Basing himself on the fact that *Lala* is a family name among the Arbëresh of Sicily and an Arvanitika toponym in Greece, Çabej argues against the Turkish origin.

A different Albanian word is the source of a term for 'grandfather' in Aromanian; the form given as *ghiuş* in Vrabie 2000: 345 and in Papahagi 1974: 614 as *gjiuşŭ* for 'grandfather,' occurring beside the inherited or nursery words *tat*, *bunic*, and *pap*, is a borrowing from Albanian *gjysh* (cf. also *aush* in Cuvata 2009).[94] Similarly, dialectal Macedonian (Steblevo, Debar region) has *gjuša*, and Mrkovići Montenegrin has vocative *đišo*, both from Albanian *gjysh*. Further, Macedonian of Gorno Papradnik (Debar region) has *kodžo* for 'grandfather,' from Turkish *koca* 'elder.' Finally, we can note BCMS *čukund[j]ed/šukund[j]ed* 'great-great-grandfather' (i.e., the father of *prad[j]ed* 'great-grandfather'), where the prefixal element is from Turkish *kökün* 'root, foundation, basis' (Škaljić 1966: s.v.). The same prefix produces *čukunbaba/šukunbaba* 'great-great-grandmother.'[95]

H. Brown 2003: 92ff. on the pragmatics of πάτερ in Homer). Aromanian also has *aféndi* 'father' (Vrabie 2000: s.v.), presumably from Greek, because of the *a-*, and not Turkish *efendi* 'sir' (though the Turkish is ultimately from Greek; see footnote 274).

93 It is worth noting that Newmark 1998: s.v. also includes various shades of meaning for 'inhabitant of Myzeqe.'

94 The Albanian form is most likely an inherited word, probably deriving from Proto-Indo-European **sū-* (**suH-*), as 'progenitor' or the like.

95 The Mrkovići datum is from Morozova 2019 (and sources cited therein). The Macedonian data are from Stoevska-Denčova 2009: 47. For *kodžo*, she feels it entered via Albanian *koçobash* 'chief elder.' However, since the Turkish form has a voiced consonant (*koc-*), as in Macedonian, whereas Albanian here has an (unexplained) voiceless < ç >, the dialectal form could have come directly

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Following on these details about '(grand)father,' we present here a sketch of further relevant kinship term loanword evidence in the Balkans. This evidence is organized by semantics, starting with the most immediate relatives and working "outward" from there, rather than by language.

#### 4.3.1.1 'Mother' (and 'Grandmother')

The sememe 'mother' in the Balkans appears to be somewhat more stable than 'father,' in the sense that it shows less evidence of borrowing.[96] There are some forms for which borrowing is plausible, but, as with 'father,' there are etymological puzzles also with 'mother' involving teasing apart nursery origins from loanword origins.

For instance, StAlb *nënë* (Geg *nânë*) could in principle be a nursery word in origin or a borrowing from Turkish *nene* 'grandmother,' a variant of *nine* 'idem'; the etymological dictionaries that mention this word (Çabej 2002; B. Demiraj 1997; Huld 1984) side with the nursery-word hypothesis,[97] and this is certainly reasonable. Turkish is not even mentioned, presumably because of the meaning difference between *nene*/*nine* and *nënë* (and the nasality in Geg); however, there is an Ancient Greek form of nursery origin, μάμμη, that means both 'mother' and 'grandmother,' so that parallels do exist for generational shifts in meaning involving female parents (and see below regarding Albanian *dadë*).[98] The situation is similar with dialectal Macedonian *nana*/*nona*/*năna*, which some see as a Turkism and others reckon as native Slavic (Stoevska-Denčova 2009: 37). In fact, however, given that the Macedonian dialectal forms are from the Debar region (Stoevska-Denčova 2009: 187), the obvious source is actually Albanian, where these various reflexes of Geg (and Common Albanian) *â* appear in the Albanian word for 'mother.' A similar ambiguity obtains for Mrkovići Montenegrin *ne$^{a}$na, nana* 'mother' (Morozova 2019). Relatedly, Greek has νενέ in the meaning 'granny, grandmother,' and this is best taken as a borrowing directly from Turkish, an account (endorsed by Andriotis 1983: s.v.) that explains the position of the stress. Similar considerations apply to the other Albanian words for 'mother,' *ëmë* (Geg *amë*) and *mëmë*, which are generally taken (so Çabej 2002; B. Demiraj 1997; and Meyer 1891) as nursery words, perhaps of considerable age, though interestingly there is at least one divergent opinion: Paşcu 1925: I, 823 takes *ëmë*/*amë* to be a borrowing from Latin *amma* 'wetnurse.'

from the Turkish without Albanian mediation. See Joseph 2016b for discussion of Albanian influence on a Greek dialect form for 'great-grandfather.'

96 Perhaps this imbalance had to do with the greater degree of access to Turkish on the part of Balkan males (see §3.0 on this gender-based differential bilingualism).

97 Çabej: "*fjalë onomatopeike*" 'onomatopoetic word,' Demiraj: "*uraltes Lallwort*" 'ancient nursery-word,' Huld: 'an archaic reduplication.'

98 And there are shifts in the meaning of kin-terms of a somewhat different nature too, e.g., Ancient Greek νύμφη meant both 'grandmother' and 'mother-in-law,' quite reasonably since both denote older generation females likely to be (or have been) part of a multi-generational and/or extended households, and, once children were born to the married couple, in identical roles vis-à-vis the grandchildren. Similar usages are found in other Balkan languages.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

This 'nurse' connection may or may not be right for *ëmë*/*amë*,[99] but it is interesting that the Aromanian word for 'mommy,' *dádă*, has parallel forms in neighboring languages with meanings that include 'nurse (for a child)': Trk/BSl/BCMS *dada* 'older sister,' Alb *dadë* 'female servant,' Trk *dada/dadı* 'child's nurse' (said to be ultimately from Persian), and Grk νταντά 'nanny.' These surrounding forms may or may not be related to Aro *dádă*, though the range of meanings for Alb *dadë* (from Newmark 1998: s.v.) is suggestive since 'mommy' is included: 'wet-nurse; pet name in baby talk for the baby's female caretaker; grandma, mommy, big sister.' Still, sources that cite them are generally noncommittal; Papahagi, for instance, notes these forms without saying specifically that any are borrowings or donor forms.

Further, Turkish *lâla*, cited above in §4.2.1, figures indirectly in this sememe, in that there is a Greek form λαλά 'grandmother'; given that the Turkish word refers to males, this Greek form is most likely a derivative within Greek from the attested masculine λαλάς 'uncle, grandfather, mentor,' which is directly from Turkish (Andriotis 1983: s.v.). The Greek pattern of masculines with a nominative in *-Vs* and a corresponding feminine in *-V*, e.g., αδερφός 'brother' ~ αδερφή 'sister,' δάσκαλος '(male) teacher' ~ δασκάλα '(female) teacher,' πατέρας 'father' ~ μητέρα 'mother,' is thus responsible for the feminine form here. The word for 'grandmother' itself can be borrowed: Slavic *baba* is probably the source of North Albanian *babë* (definite: *baba*) 'grandmother, aunt, form of address to old women' (Mann 1948: s.v.; cf. Curtis 2012: 79).

Finally, there is one interesting use for Alb *ëmë*/*amë* that does show borrowing, but from Albanian into Bulgarian. Çabej 1996: 118 reports that within Albanian, *amë* acquired an initial *t-* from a preposed particle of concord (*të*), possibly as an accusative, *t(ë) amë*, or via a resegmentation involving *jot* 'your,' where the *-t* is etymologically connected to the root for 'you,' and this new form *tamë* took on the meaning 'fountainhead, source' (i.e., presumably a metaphor, the 'mother of the waters'). Further, Çabej continues, *tamë* has been incorporated into the secret language of Bulgarian-speaking bricklayers in the village of Smolsko in the Pirdop region of Bulgaria as *tama* 'mother'(Kănčev 1956: 402).[100]

### 4.3.1.2 'Brother'

Turkish *ağa* 'master, patron' (older and dialectal *aga*, which is its shape in the Balkan languages) is also a provincialism meaning 'older brother' (StTrk *ağabey*) and was used colloquially in this meaning in Bulgarian (Grannes et al. 2002: s.v.).

99 We must remember that etymologizing is not an enterprise where one takes a vote and decides the origin of a word according to the will of the majority. Still, short of working out the details for oneself in each case, one can only turn to authoritative sources for guidance, as we do here.

100 See §4.4.3 for more on secret languages in the Balkans. These speakers might have been from Macedonia, which provided migrant labor, especially in masonry, throughout the Ottoman Empire, and thus in modern terms, the form is probably via Macedonian.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

The versatile and seemingly ubiquitous (in these sections at least) Albanian term *lalë* offers another case. For Albanian, Meyer 1891: 236 states that this word, which is a borrowing from Turkish *lâla* (see §4.3.1.1), has the meaning 'elder brother' in Kavaja and in Myzeqe in general; further, Arbëresh of the Bova region has *leḍḍé* for 'brother' (and a derivative from that, *leḍḍá*, for 'sister'), which, according to Meyer, is from this word.

Aromanian *baci* is given in Vrabie 2000: 173 under the lemma for 'brother,' for which the regular term is *fráte* (from Latin *frater*). The word *baci* is labeled as a 'term of respect for an elerly [*sic*] brother.' It is thus not a primary kinship term, despite the fraternal meaning, but it is connected to kinship terminology. A similar word, *bac[ë]*, is a Geg regionalism meaning 'older brother, father, or father's brother' (Newmark 1998: s.v.). The term is used regularly for 'older brother' in Kosovo. The source is the identical *bac* in South Slavic (cf. BCMS *bac*, Mac *bate*, with an affective affrication of *t* to *ts*; cf. example (5.31b) in §5.7).[101]

### 4.3.1.3 'Sister'/'Daughter'

Aromanian *dódă* for 'older sister,' is given in Papahagi 1974: 497 as being of unknown etymology. This form means not just 'older sister' but also 'aunt' or 'grandmother' or 'wet nurse.' While it could well be merely an Aromanian-internally derived nursery form, it does show some formal and semantic connections with words in neighboring languages that should not be ignored; these include Trk/BSl/BCMS *dada* 'older sister,' Alb *dadë* 'big sister; grandma' (see §4.3.1.1). The polysemy of the Albanian word is striking when compared with the parallel polysemy of the Aromanian, making a borrowing hypothesis appealing, even if the direction of the borrowing cannot be definitively determined, and even if some internal influence, perhaps a nursery-related effect, was responsible for the vocalism of the initial syllable in *dódă* (which is also attested in Moldavian). Similarly, Romanian, based on Meyer 1891: 236, has *lele* in the meaning 'older sister,' seemingly connected with the Turkish *lâla* form cited above (as a feminine derivative of a presumed masculine borrowed form), but taken, more compellingly, by BER III: 357 as a loanword from Bulgarian *lelja* 'aunt' (on which see below, §4.3.1.5), despite the difference in meaning.

Further, along these same semantic lines, there are words that can mean 'daughter' – as well as 'young girl,' so that they are not necessarily primary kinship terms – that are common across several of the languages and for which a borrowing origin is generally accepted. Albanian has *çupë* in the meaning (from Newmark 1998: 148) 'girl, lass; little girl, daughter; unmarried woman, maiden,' and it is borrowed into southwestern Macedonian as *čupa* 'girl' and Aromanian *čiup* 'small child' (Papahagi 1974: 451) and the meaning of the related *čiúpră* 'daughter,' a word that appears to come from an

101 Words denoting 'brother' or 'older brother' were affected as to their consonantism by a nursery word throughout Slavic, e.g., Russ *batja*, Cz *bat'a*, Blg *baj*, among others. Cf. also Russ *batjuška* 'little father,' an epithet of the Czar.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Albanian collective in *-r- *çup[ë]ra,* cf. *çupëri* 'girlhood; girls taken collectively, the world of girls' (Newmark 1998: 148).[102] The Macedonian form can also be used for 'daughter' (as, however, can the Slavic-derived word for 'girl', *devojka*).[103] Similarly, Greek shows both τσούπα and τσούπρα, in the meaning 'young girl, daughter,' also clear loanwords from Albanian (or via Macedonian), again with the occurrence of the Albanian collective marker *-r-* (here -ρ-) as a telltale sign of borrowing.[104] Turning to a different word with a similar meaning, Albanian *bijë* 'daughter' is the source of (dialectal) Macedonian, Kosovo Serbian, and Montenegrin *bija*.

#### 4.3.1.4 'Uncle'

There are some borrowings to be noted among words for 'uncle' in the Balkan languages. First, the Turkish word *dayı* 'maternal uncle; mother's brother' is clearly the source of Alb *dajë*, which has the same specific meaning the Turkish word has. This holds as well for Balkan Slavic, where Macedonian has derivatives (diminutives of endearment) *dajče* and *dajko*, and Bulgarian has *daija* and the derivative *dajčo*, occurring alongside the native Slavic *vujko* for 'mother's brother, maternal uncle.'[105] Also, in some Bulgarian and Macedonian dialects, especially among Muslims, *dajo* occurs for 'maternal uncle' (BER I: 314).[106] Greek has νταής 'bully,' cf. StTrk *kabadayı* 'bully' (StTrk *kaba* 'rough, coarse, crude, vulgar'; cf. also BCMS *da[h]ija* 'renegade janissary, tyrant'). Further, Meglenoromanian has *daiã*, Romani has *dajos*, and both Bulgarian and Macedonian show, dialectally (Stoevska-Denčova 2009: 94), *kalèko* and *kal'eko*/*kăl'eko* for 'aunt's husband,' from Greek καλο- 'good' (quite likely from the vocative καλέ '(my) good (man)!') with a diminutive suffix. Finally,

102 *Čiúpră* is found in both Papahagi 1974 and Vrabie 2000 in addition to *hilji* (Cuvata 2006: s.v., < Lat *filia*).

103 The normal word for 'daughter' in Macedonian, however, is inherited *kjerka*.

104 The source of Albanian *çupë* does not necessarily pertain directly to intimate loans within the sprachbund, but it is interesting nonetheless. It is best considered Slavic (so Meyer 1891: 450; Skok 1971: 342–343; more cautiously Çabej 1987: 144–145), since there is a Serbian form *čupa* 'tuft (of hair); woman with uncombed hair' with cognates elsewhere in Slavic; the meaning shift would be from 'long-haired (one)' to 'girl' (as the prototypical long-haired one in the family). Çabej argues that since Albanian is the source for the Macedonian, Aromanian, and Greek forms, it is possible that in Albanian a local word (*një fjalë vendi*) was mixed (*u përzie*) with Slavic. Interestingly, there are no hair-related meanings given in Newmark 1998: 148 for *çupë*, but Meyer 1891: 450, writing more than a century earlier, gives, in addition to the meaning *Mädchen* 'maiden, girl,' the definition *langes Kopfhaar* ('long head-hair'). Meyer writes that *çufkë* 'tuft, etc.' belongs here and not with Italian *ciuffo* 'tuft.' However, given that Albanian *xhufkë* 'tuft, tassel, topknot' is a highly productive base (cf. *xhufkor* 'fringed,' *xhufkon* 'to tassel [of plants]', etc.) while *çufkë* is merely a variant of *xhufkë*, it seems likely that the usual Albanian form is from dialectal (southern) Italian, whence it became productive, while the variant is later from northern Italian. The *-kë* here is the Slavic suffix discussed in §4.2.2.3.

105 Macedonian also has the formation *kjose-dai* (Jašar-Nasteva 1987: 88), involving a derivation from *kjose* 'beardless man,' a borrowing from Turkish *köse* 'with little or no beard' plus *dai*, from WRT *dayi* 'uncle.'

106 And the Turkish word is found in a variety of forms in BCMS: *dája*, *dájka*, *dájko*, and *dajo*, as well as *daidža* where *-dža* reflects a Turkish diminutive suffix *-ca* (Škaljić 1966: s.v.).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Albanian (Geg) has *bac[ë]* 'uncle, older brother, etc.,' from Slavic (cf. Curtis 2012: 79).

The occurrence of diminutives with this word is not surprising, given that they can fall under the realm of intimate kin terms. Indeed, within Turkish, a diminutive suffix *-ca* seems to figure in the derivation of the widespread word for 'father's brother; paternal uncle,' *amca*, composed of a word of learnèd usage from Arabic (so Redhouse 1968: 55–56) *am* (Arabic *'amm*) 'paternal uncle,' with a diminutivizing *-ca* (usually found on adjectives).[107] Turkish *amca* is important in the Balkans as it is the source of Bulgarian *amudža* and the most likely source of the Albanian words for 'paternal uncle,' *xhaxha* and *axhë*, and of Macedonian *adžo* (found in Tetovo and elsewhere) for 'father's brother; older man,' as well as BCMS *adža, amidža, adžo* (VOC), 'idem,' which occur alongside the native Slavic form *striko*.

Regarding the Albanian, only Meyer 1891: 79–80, among the various Albanian etymological resources, says anything about *xhaxha*, and he is noncommittal, citing only OCS *dědъ* 'grandfather' and Russ *djadja* 'uncle,' for which latter CoSl **dēdŭ* is the etymon, but for which the reflex *ja* from **ē* results from an East Slavic assimilation of a front nasal and is not related to anything West Balkan (cf. Vasmer 1986–1987: s.v.). Although Meyer writes that "the word may be present even in the Slavic Balkan languages," the word in question woud have to be *ded* or *djed*. It is not straightforward to derive either *xhaxha* or *axhë* from *amca*, but the key may be the form *xha*, a "respectful title used in addressing an older man by his first name" (Newmark 1998: 946). This form conceivably was abstracted out of *amca*, which can also be used as a "term of address to an older man" (Redhouse 1968: 56), and then reduplicated within Albanian, perhaps as a nursery effect, to give *xhaxha*. Still, even under this account, contact with Turkish was involved in the derivation of *xhaxha*. *Axhë* and *adžo*, then, are either from *amca* via a phonetic reduction in the borrowing of the otherwise unusual cluster [mdž], or else abstracted out of *xhaxha*, as if it were segmented *xh-axha*.[108] The Macedonian form could thus be an Albanian loanword, though with the *-o* as the result of it being drawn into the morphology of hypocoristic kin terms (cf. *striko*, *tetko*).

The Turkish word *çiçe* 'aunt' seems to be the connected to Slavic *čičo* 'uncle' – thus with a now-familiar change of gender: BSl and BRo have *čičo*, *čičko*, *čiča*, *čika* (BCMS) / *cică*, *cicio* (regional), *čiča*, *tsitsă*, respectively, cf. also Alb *çeço* 'daddy, eldest brother.' These could in principle be independent, Slavic-based creations, and Stoevska-Denčova 2009: 86–87, though inclined towards the Turkish etymology, equivocates as to the source (Skok 1971 does not give a source, reflecting uncertainty in the relevant literature).

Albanian offers another case of a loanword for 'uncle' in the general term *ungj*, as this is a borrowing from Latin *avunculus*. The contraction of *awu-* to *u* shows

107 There are variants, labeled as "provincial" in Redhouse 1968: s.vv. *amu* and *amuca*. This latter is surely the source, via some sort of clipping and/or resegmentation, of Albanian *mixhë* 'uncle,' and, either directly or via Albanian, of Macedonian *midžo* (with *-o* as in *striko* and *tetko*).

108 This cluster does occur in Macedonian in some words of Turkish origin, e.g., *džamdžija* 'glazier.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

a regular sound change of (post-)Roman-era Albanian, and the *-ngj* derives from the syncopated *-ncl-* with regular voicing induced by the nasal and the expected *-gj-* outcome of a *-gl-* sequence.

In Romanian, besides inherited *unchi*, also from Latin *avunculus*, the form *nene* and variant *nea* and diminutive *neică* all mean 'uncle' as a term of respect rather than kinship. The source is Slavic (cf. Skok 1972: s.v. *naja*, BER IV: s.vv. *nena, nenjo,* Vasmer 1986–1987: s.v. *njanja*). Aromanian, however, has *lálă* for 'uncle' (Vrabie 2000: s.v.), possibly a loanword although its origin is hardly certain. Papahagi suggests a connection with Latin *lalla* 'lullaby' (not attested as such in Latin but presumably based on the verb *lallo* 'sing a lullaby') but that seems rather far-fetched as a source for 'uncle.' A better starting point for the Aromanian is Albanian *lalë* 'elder brother; (paternal) uncle; godfather,' itself probably a borrowing from Turkish *lâla* 'manservant assigned to the care of a child' (ultimately from Persian). This Turkish word is also the likely source of a nineteenth-century Macedonian form *lală* for 'uncle' cited by Meyer 1891: 236 but not current through much of the twentieth century (see §§4.3.1, 4.3.1.1, and 4.3.1.2 for other ways that *lâla* has had an impact on Balkan kinship terminology).

### 4.3.1.5 'Aunt'

A word for 'aunt' has been discussed, in §4.3.1.3 above, with regard to Aromanian *dódă*, meaning 'older sister' but also 'aunt' (and 'grandmother'); this word seems best taken to be a loanword from Albanian, although the Albanian source does not show the 'aunt' meaning.

There are, however, clear cases of loanwords for 'aunt' in the Balkans. Aromanian offers two other words for 'aunt' that are likely borrowings: *tétă*, taken by Papahagi 1974: 1176 to be from Bulgarian (or Balkan Slavic more generally) *teta*, and *ţáţă*, from Greek τσάτσα. The Greek may reflect a reduplication within Greek of a form based on θεία/θειά 'aunt,'[109] but it is hard to separate it from the Turkish slang term *çaça* 'woman who keeps a brothel.' Interestingly, though, *çaça* is said (Redhouse 1968: 235; Tietze 2002: s.v.) to be from the Greek τσατσά for 'old woman' (derived by an accent shift from τσάτσα 'aunt(y)'), which can also have the 'brothel'-related meaning of *çaça*;[110] thus the directionality of the borrowing may not be clear. For Bulgarian, Gerov 1895–1908: s.v. gives *čičjá* for 'father's brother,' and *číčja* for 'father's brother's wife.' It is possible that Turkish *çiçe* 'aunt' played a role here, although the accentuation makes this hypothesis problematic.

109 So Andriotis 1983: s.v., following Filindas (cited as "Γλωσσογνωσία 1.149" (cf. also Tietze 2002: s.v.)), though it does require an irregular (but not necessarily unparalleled) sound change if τσι(γ) γάνος 'Gypsy' is from α-θίγγανος 'untouchable' rather than from Turkic *čiɣan* (Matras 2011: 257; see §4.3.9.4 and footnote 310 below, and Chapter 5, footnote 210); Filipova-Bajrova 1969: 882 suggests, however, that it derives from Bulgarian *tsitsa* 'aunt.' Note that θειά is from earlier θεία by regular sound changes: [θía] > [θiá] > [θjá].

110 So Charalambakis 2014: s.v.; LKN: s.v.; Babiniotis 1998: s.v.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Turkish itself distinguishes in its kinship terms between 'maternal aunt (mother's sister),' *teyze*, and 'paternal aunt (father's sister),' *hala*. These words – and the associated semantic distinctions – were borrowed into Albanian as *teze* and *hallë*, occurring alongside the more general terms *emtë* 'aunt (paternal or maternal),' itself a loanword from Latin *amita* 'paternal aunt,' and *teto* 'aunt,' a likely borrowing from Macedonian, though a nursery origin cannot be ruled out. Dialectally, Macedonian and BCMS have *ala* from Turkish *hala* (Stoevska-Denčova 2009: 178, citing Jašar-Nasteva 2001; Morozova 2019), while Kratovo and Tetovo Macedonian have *teza*, though not *hala*. For BCMS, Bjeletić (1995: 208–209), cited in Morozova 2019, notes that both *ala* and *teza/teze* mean simply 'aunt' in most BCMS dialects where they occur, and it is only in Mrkovići and the Catholic village Janjevo in Kosovo that father's and mother's side are distinguished. Similarly, Bulgarian has *tejza*, *teze* (dialectally *tize*), from Turkish, though these are now considered to be obsolete; interestingly, Bulgarian *ale*, *hala* are cited as 'maternal (sic!) aunt' in Morozova 2019, as are *teza*, *teze*.[111] Many dialects of Romani in the Balkans have *tetka*, and Macedonian Arli and Kosovo Bugurdži also have *teza* in addition to *tetka*. Native Romani is *bibi*, a form also widely used in the Balkans.[112]

Albanian is a source of words for 'aunt' in various Macedonian dialects, as documented in Stoevska-Denčova 2009: 90–91. In various villages in the Debar region, *džidža* and *džedža* occur, apparently borrowed from, or better, based somehow on, Albanian *xhaxhkë* 'aunt.' Similarly, in Slimnica (Grk Trílofos) in the Kostur (Grk Kastoria) region, Kunovo in the Gostivar region, and Suho in the area around Thessaloniki, *nana*, *nane*, and *načа* occur, respectively, based on Albanian *nënë* 'mother.'

A familiar etymological puzzle also arises here. Bulgarian has a form *lelja* for 'aunt' that would seem to have something to do with Turkish *lâla*, perhaps involving a feminine derivative of a(n unattested) masculine form taken directly from the Turkish. However, Meyer 1891: 236, writing that *lelja* should be separated from the *lâl*-related forms, urges caution here, appropriately enough since, as BER III: 356–357 shows, forms related to *lelja* are to be found all over Slavic and Baltic as well, revealing it to be a Balto-Slavic lexeme of long standing.

The *lâl-* word in Albanian, *lalë*, which otherwise has male meanings ('young father; elder brother,' etc.) does show a gender-shifted meaning to 'aunt' in Geg, according to Meyer *ibid.*, where the form is *jajë* (marked as nonstandard by Newmark 1998: 334). The exact mechanism for this shift is not entirely clear but may involve an internal derivational process.[113] Meyer 1891: 91 notes that this

111 Grannes et al. 2002, for instance, make no mention of *hala*, which is however found in BCMS (Škaljić 1966: s.v.). In the context of borrowed kinship terms, it can be noted that Turkish *hala* 'mother's sister' itself is a borrowing from Arabic *ḫāle* (cf. older Turkish *hâle*, Ayverdi & Topaloğlu 2006: s.v.).

112 For the sake of completeness, we mention here that for 'aunt,' Romanian has *mătușă*, taken by some to be based on Latin *amita* 'aunt' with a suffix *-uș-* but by others as a borrowing from Ukrainian *matuša*.

113 Given that there can be generational shifts in the meaning of some kin terms (as seen in §4.3.1.3 regarding Albanian *dadë* 'big sister, grandma' or §4.3.1.4 regarding Albanian *bacë* 'uncle; older

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

form shows an interesting development in the Berat (Tosk) dialect, where the form is *thjajë*; Meyer accounts for the unusual form by appealing to influence from Greek θεία ([θjá]}, giving a form that is a phonological loanblend or hybrid.

Finally, Turkish has figured heavily in the borrowed kinship terminology documented in the preceding sections, but as the donor language in case after case. There are, of course, kin terms in the Balkans borrowed from donor languages other than Turkish, as shown by various examples throughout. Still, one such case deserves mention here involving Turkish as the recipient language. Turkish as spoken in North Macedonia has borrowed the Macedonian word *tetko* 'aunty,' for use as a vocative. Many dialects of Romani also have *tetka/tetko*. Thus donor and recipient language in the Balkans are not predetermined roles; rather they depend on the local social circumstances.

### 4.3.1.6 'In-Laws'

As with other kinship terms, words for various in-laws in the Balkans also yield instances of loanwords. For instance, Bulgarian and eastern Macedonian have *baldəza* (with a variant *baldəzka*, standard Mac *baldaza*); BCMS (in Muslim contexts) has *balduza* (Škaljić 1966: s.v.) or *balgaza* (Morozova 2019) for 'wife's sister,' and Macedonian Arli Romani has *baliska* (Halwachs et al. 2007) thus a type of sister-in-law, a borrowing from Turkish *baldız* 'sister-in-law, wife's sister.' Kratovo and Tetovo Macedonian *gelin*, Montenegrin and Shkodran Geg *gjelinë,* and Mrkovići Montenegrin *đelina*, as well as Pomak Bulgarian *gelina* (Morozova 2019: 333), all refer to both 'bride' and 'daughter-in-law,' from Turkish *gelin* 'bride; daughter-in-law' (absent from Škaljić 1966). Dialectal Macedonian *nusa* (in Gora, both *nusa* and *nuse*) from Albanian *nuse* 'bride; daughter-in-law' also occurs in parts of southern Montenegro (Morozova 2019). In Kratovo one finds *jenga* 'sister-in-law,' from Turkish *yenge* '(woman's) sister-in-law; aunt-in-law' ('husband's brother's wife'); Škaljić 1966 gives the following forms for BCMS: *[j]enđa, [j]enga* (with both *ë* and *é*), *[j]enđija, [j]engija* (with *è*) with meanings varying from 'husband's brother's wife' to the equivalent of 'maid of honor at a wedding' (who is usually a female relative of the equivalent of best man *kum, stari svat*). For Bulgarian, Grannes et al. 2002: s.v. gives *engé* 'father's brother's wife,' and Jašar-Nasteva 2001: 88 notes *jengja* for 'father's brother's wife' in Kratovo Macedonian.

Dialectally in Macedonian forms such as *kain* and *kaim* (Trk *kayın*) occur for 'wife's brother' and BCMS has *ka[j]in* (Škaljić 1966: s.v.) as well as diminutive *kainče*. In Bulgarian, there are the variants *kaínče, kaínčo* 'idem' and also *kaína,*

brother'), one might be tempted to think of some connection between Geg *jajë* and Greek γιαγιά, phonetically [jajá]), but the fact that within Albanian it is restricted to Geg makes this unlikely. Andriotis 1983: s.v. says merely that γιαγιά is a nursery word, a view taken also in Dangitsis 1978: s.v. and Babiniotis 2010: s.v.; Floros 1980: s.v. suggests a Turkish origin for this Greek word, but there does not seem to be a suitable Turkish word in any major dictionary (Redhouse 1968; Akalın & Toparlı 2005; Ayverdi & Topaloğlu 2006; TDK 1963–1977; TDK 1963–1982).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

*kájna*, and *kaína*, and *kaínla* (<? *kayin abla* = *kayın* + *abla* 'sister'[?]) 'husband's brother's wife' (see below for more on 'brother-in-law').

Greek is involved here insofar as forms based on πεθερός 'father-in-law' occur dialectally in Macedonian (Stoevska-Denčova 2009: 115): *p'efir* in Negovan (Grk Ksilópoli), near Thessaloniki, with *f* for *θ* and an *-i-* that reflects the usual northern Greek pronunciation; *pehjar* in the Sérres region, with *h* for *f*; and *peăr*, also in the Sérres area, with (regular) loss of /h/.

Perhaps the most widely distributed Turkish kin term in the Balkans is the word for 'husband of one's wife's sister; brother-in-law' (and derivatives from it). The Turkish word is *bacanak* 'brother-in-law (wife's sister's husband),' and it has yielded Mac, Blg, and BCMS *badžanak*, Alb *baxhanak*, Aro *baginac, bãginac*, Megl *baginac, bãgiãnac*, and Grk μπατζανάκης (cf. Morozova 2019). The meaning in Albanian, Macedonian, and South Danubian Balkan Romance (SDBR) retains the specificity of the Turkish form, whereas in Bulgarian and Greek the meaning has been generalized (so BER I: s.v.) to cover 'sister's husband' as well (thus competing, for Bulgarian, with the inherited Common Slavic term: OCS *šurinъ, šurъ*, Mac *šura*, Blg *šurej*, Russ *šurin*, etc.). The standard Turkish form is as given above, but there is a variant form given in Redhouse 1968: 116 as *bacınak* that might explain more directly the SDBR forms with a medial high front vowel. Similarly, the variants with schwa in the first syllable, could represent vowel reduction or, perhaps, dialectal Turkish.

The "success" of the spread of this foreign word in some of the Balkan languages could be a consequence of there being no inherited Indo-European term for this particular relation; the Indo-European form continued in AGrk δαήρ, Lat *levir*, Mac/Blg *dever*, etc., meant 'brother-in-law,' but, based on the precise meaning for these cognate forms, it was 'brother-in-law' as 'husband's brother,' and thus different from the Turkish form.[114] Balkan Slavic, however, already had a specific native term in place by the time of contact with Turkish. One can speculate, however, that during the Ottoman period this particular affinal relationship took hold in promoting networks of solidarity, a role that *badžanak* continues to have in North Macedonia to this day.[115]

Finally, some Balkan terms for other in-laws are derivatives of these or reflect other borrowed kinship words. For instance, 'sister-in-law' in Aromanian is *bãginacã*, a feminine form derived from the masculine *bãginac*. And, in Bulgarian (Grannes 1996: 164) and dialectal Macedonian (Stoevska-Denčova 2009: 128), *babalăk* occurs for 'father-in-law,' taken directly from the Turkish *babalık* 'fatherhood; stepfather; father-in-law; adoptive father' (Redhouse 1968: 115).

114 Interestingly, Slavic *dever* is borrowed into Albanian, but not necessarily as a kinship term, rather as 'boy who leads the bride's horse; paranymph' (who may or may not be a brother of the groom, depending on circumstances), thus in the same general semantic sphere as kinship terms, being connected to ceremonies which create some kinship relations.

115 Thus, for example, the term occurred in numerous headlines in internet news sources in Macedonian and Albanian in 2019 referring to accusations of nepotism.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.1.7 Larger Kinship Units

The data in this section are taken from Sobolev 2006: 14 and are intended to be illustrative rather than exhaustive. In English, the term *family* can be *nuclear*, *extended*, or in reference to a larger group of kin. Terms such as *clan* and *tribe* tend to be ethnographic or informal, often with humorous connotations in the latter case, while *lineage* is a technical term. In the Balkans, as in most of the world, degrees of group familial relationship were, and in some places still are, denoted by specific terms. Thus, for example, Albanian *fis* is still a significant larger kinship unit in many regions. The Slavic equivalents are *pleme* and *rod.* Of interest here is the fact that in the Aromanian, Greek, and some of the Balkan Slavic dialects documented in Sobolev 2006: 14, Turkisms either coexist with or are the sole expressions of such larger family units. Thus in Turia (Grk Krania) Aromanian, in the Pindus region, the Turkism *soy* co-exists with *rădătsină* and *rizã* (this last a Hellenism meaning *root*). *Soy* is also used in both the northern and southern Greek points (Erátyra and Kastélli, respectively), and the other term, *damar/ndamari* is likewise from Turkish. In northeastern and Rhodopian Bulgarian (Ravna and Gela, respectively), the Turkism *džins* (Trk *cins*) co-exists with native *rod* (although the Turkism is now archaic in Ravna).

One other Balkanism involving larger kinship units, based on Sobolev 2006: 14, is worthy of note here. The data for Albanian give only *fis* for the kinship unit equivalent to Slavic *rod*. In Albanian, *vllazni* (Geg), a collective derived from *vlla* 'brother' (StAlb *vëlla*), is a sub-unit of *fis* roughly translatable as 'clan.' The Slavic etymological equivalent of *vllazni*, however, *bratstvo*, is the term used for the larger descent group in Zavala (Montenegro), while *rod* is normally used for the bride's family. Given that this usage occurs precisely in Montenegro, we can posit Albanian semantic influence.

### 4.3.1.8 Fictive Kinship

As K. Brown 2005: 45 writes: "Anthropologists working in [Serbia, Macedonia, Greece, Albania, and Bulgaria] have documented the importance of what they term *fictive kinship*, whereby people unrelated by blood [or marriage –VAF/BDJ] forge bonds that are enduring and sacred." The example Brown cites is that which is termed *kumstvo* 'godfatherhood' in Balkan Slavic. The term *kum* can be translated 'best man [at a wedding]' but also as 'godfather,' as he takes on responsibiities for the children resulting from the wedding.

Late Latin (Balkan Latin?) *compater* and *commater* 'co-father/co-mother' seem to be the ultimate source of BSl *kum/a*, (whence Arli Rmi *kum*), Alb *kumbar/ë*, Grk κουμπάρ-ος/α (whence *kirvo*, etc. in many Romani dialects, Boretzky & Igla 1994: s.v.), Rmn *cumar*, and Aro *cumbar/ă*, in part through OCS *kъmotrъ* and in part through Venetian *compare*, though the specific paths of diffusion within the Balkans are probably lost to history.[116] The synonymous *kalitata*, found dialectally

116 Babiniotis 1998: s.v. gives Venetian *compare* as the source of the Greek.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

in Macedonian (Stoevska-Denčova 2009: 166), appears to derive from Greek καλή τάττα (lit., 'good aunt'), probably influenced by Macedonian *tate* 'daddy.' The Albanian form *kumbara* has also been borrowed into the BCMS of Kosovo (Morozova 2019 and sources cited therein). Ultimately, Latin *nonnus* 'monk,' *nonna* 'nun, childcarer' (cf. Itl *nonno* 'grandfather,' *nonna* 'grandmother') is the source of Grk νουν(ν)ός, dialectal νούννος 'best man and subsequently godparent of first child, one who holds a child at baptism, one who gives child first haircut, etc.' (also νονός, νονά), also Alb *nun* 'godfather, best man' and *nunë* 'godmother, mother of baby getting its first haircut from its godfather,' and from one of these sources, Mac *nun(ko)/nunka* 'godfather/mother,' Blg (Svilengrad region) *nunjo/nuna,* Aro *nun/nună* 'idem' (BER IV: s.v.).

The abovementioned Latinate word complexes relate to life cycle events associated with marriage and birth (see §4.3.11), but another socially important fictive kinship relationship in the Balkans was that of blood-brotherhood or -sisterhood. Such fictive kin relations were an additional way to increase solidarity in interpersonal relations. In general, the various Balkan languages have terms of native origin for this culturally shared institution, e.g., for 'blood-brotherhood' SSl *pobratimstvo,* Grk αδελφοποιΐα, Aro *fărtătsilje, fărtătliche* (with Turkish *-lik*), Megl *fărtățília,* Rmn *frăție de cruce* ('of the cross'), StAlb *vëllami.* It is therefore of interest to note the distribution of borrowed terms for this institution in Albanian: for 'blood-brother' (StAlb *vëllam*), the Slavic *pobratim* is used throughout Kosovo and Northern Albania as far south as the Mat River as well as in Zajaz (Kičevo/Kërçova), while most of Central and Southern Geg as well as scattered Tosk points as far south as Muzhakat (Grk Mouzakéïka) in Çamëri/Epirus use the Turkism *byrazer/burazer/birazer* (cf. StTrk *birader*; Gjinari 2008: Map 265).[117] We can also note here that the Macedonian dialect of Shulin (Lower Prespa region, Albania) has *kušer* for 'cousin' (Stoevska-Denčova 2009: 82), a clear borrowing from Albanian *kushëri* 'idem' (itself a borrowing from Latin), which, while denoting a genealogical rather than a fictive kin relationship, pertains to a similar social function of extending familial solidarity. Similarly, *adžovci* 'paternal cousins' (from the Turkism *adžo* 'paternal uncle' with Slavic suffixation) occurs among Muslim speakers of BCMS (Morozova 2019).

Another type of fictive kin relationship is seen in the shift of meaning from Grk παραμάνα 'wetnurse' as the source of the dialectal Macedonian hybrid form *paramajka* 'stepmother' (Stoevska-Denčova 2009: 142).[118] Here the cultural connection is that understood by the term *milk-mother*, a concept present in both Islam and Eastern Orthodox Christianity.

117 Cf. also Macedonian Arli Rmi *pobratimluko* 'friendship' (ROMLEX), on the one hand, and the fact that Turkish *dost* 'friend' is used for 'blood brother' in Hogosht, Kosovo (Gjinari 2008: Map 265), on the other.

118 Aromanian has *paramanã*, from Greek, in the original meaning of 'wetnurse.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.1.9 Miscellaneous

There are other Balkan instances of borrowing involving kinship-related terms, beyond what has been given. As they do not have a particular unifying theme, they are treated here just as the miscellaneous occurrences that they are.

For instance, there are some nonkinship-related Balkan formations that derive from Turkish kinship terms. The expression discussed above in §4.2.2.6.2 that is found in Greek as αναντάμ μπαμπαντάμ (also αναντάμ παπαντάμ), and in Albanian as *denbabaden*, meaning 'in or since the distant past' but built on 'mother' (*ana*/*anne*) and 'father' (*baba*), is one such case. Another is Bulgarian *babanlăk*, cited in Grannes 1996: 164 as meaning "*passé éloigné*" 'distant past.' Grannes derives it via assimilation from a putative Turkish *babam-lık*, which, like *anadan babadan*, does not occur in authoritative Turkish lexical sources (Redhouse 1968; Akalın & Toparlı 2005; Ayverdi & Topaloğlu 2006; TDK 1963–1977; TDK 1963–1982).[119] Here *babam* is 'my father' and *-lık* the Turkish abstract noun-forming suffix, so that the sense is originally "*de temps de mon père*" 'from the time of my father.' He notes, though, that *baban* occurs dialectally in Bulgarian (BER I: s.v.) in the meaning 'papa,' presumably from Turkish *babam* with final *-m* becoming *-n*; thus, *babanlăk* could be a Bulgarian creation, since the suffix *-lIK* was borrowed into Bulgarian and is quite productive (see §4.2.2 above). Still, as argued in §4.2.2.6.2, the absence of such expressions from contemporary Turkish lexical resources need not be decisive here since they may well be dialectal or colloquial phrases of a hundred or more years ago.[120]

Finally, as a somewhat secondary use of a term for a kin-determined relationship, we note Albanian *bir* 'son' occurs in Greek folk songs and folk poetry of the Greek communities in Southern Albania; in those works, *bir* refers specifically to the son of an *aga* 'lord.'[121] The word in this case is borrowed but in a highly specialized context that is related to the kin use but removed from the immediacy of close kinship. A different, and less specialized, use of borrowed *bir* (from Albanian) occurs in dialectal Macedonian (Stoevska-Denčova 2009: 58), specifically Nestram (Grk Nestório), Kostur (Grk Kastoria) region), where it is used alongside of *sin*, the native word for 'son.' (See also §§4.3.4.2.2, 4.3.8.)

### 4.3.1.10 Summary Regarding Kinship Terms

The facts about these kinship terms are interesting in their own right. There is an intriguing versatility in semantics for some, especially the words related to

119 Turkish *babalık* means 'fatherhood, stepfather, father-in-law, simple old man' (Redhouse 1968: s.v.). Regarding *anadan, babadan*, as noted in §4.2.2.6.2, the expression is uncommon in modern Turkish and is generally associated with rural or folksy contexts.

120 See also §5.4.1.1, footnote 32, for another similar example, dialectal Greek [baldyrs].

121 For instance, μπιρ is found in a popular song from southern Albania dated to the eighteenth or nineteenth century, "Όσα δέντρα έχει τ΄αλώνι" ('As many trees as the yard has,' lit., 'as-many-as trees has the threshing.floor'), published in TES 2006: 10.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Turkish *lâla*. Moreover, the fact that so many loanwords are detectable in this general semantic domain is striking, and the import of these kinship loans should be clear. Their occurrence is consistent with everything that is known about the intimate and intense contact situation in the Balkans, especially involving Turkish, coupled with widespread bi- and multilingualism.

Examples of borrowings involving kinship terms can be found even under circumstances that appear at first to be quite different from those in the Balkans. The entry of words like *aunt*, *cousin*, and *uncle* into English, from French,[122] provides a ready case in a familiar language, probably motivated, as Matras 2009: 170 puts it, by the fact that "the use of French words for family relations . . . [was] fashionable in Medieval English due to . . . an association with the terms used by the French-speaking social elite." The social mixing between English and French speakers in the Middle English period might be characterized by some as sustained and intimate contact,[123] but interestingly, as Matras emphasizes, "this fashion was not extended to closer kin"; this, he suggests, shows "a reluctance on the part of speakers to compromise certain familiar, intimate terms of everyday life," indicating a difference with the more intimate and intense Balkan situation. The chiefly British English use of Latin *mater* and *pater* for 'mother' and 'father' respectively, especially by schoolboys and sometimes facetiously (as some dictionaries indicate; see also footnote 88 and note the Turkish Persianism *peder* 'father'), may thus be a better example of the borrowing of kin terms in a context of nonintimate contact, but there the special relationship of Latin to upperclass British English speakers may have played a role (like that of Persian to Ottoman Turkish), in a Hockettian prestige-related way; alternatively, when used facetiously, we should recall the insight implicit in Weinreich 1968: passim about bilingualism extending a speaker's expressive range.[124]

Still, when coupled with other indications of borrowing based on everyday conversational interactions, such as the stylistic lowering of the Turkish loans

122 Ultimately from Latin *amita* 'paternal aunt,' *consobrinus* 'mother's sister's child,' and *avunculus* 'mother's brother,' respectively.

123 Whether English and French together therefore could constitute a sprachbund is a different question, one that might be answered affirmatively (see §3.4.2.1 on two-language sprachbunds) although the single result, modern English, looks quite different from the mutual maintenance of multilingualism of the Balkans.

124 Although he never states this – to us, key – insight in precisely these terms, Weinreich hints at it in several places in his classic work. For instance, on p. 34, he observes that as "a mechanism for the reinforcement of expression, the transfer of morphemes naturally flourishes where affective categories are concerned," by which categories he means "diminutive and endearing affixes." Further, on p. 58, he notes that a "language can also satisfy its ever-present need for euphemisms and slangy 'cacophemisms' by borrowing," and continues (p. 59) by writing that while "the unilingual depends, in replenishing his vocabulary, on indigenous lexical material and whatever loanwords may happen to be transmitted to him, the bilingual has the other language as a constantly available source of lexical innovations." And, on p. 31, footnote 5, Weinreich explicitly refers to the deliberate facetious use of material from another language.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

seen here, assigning these instances of borrowing in the Balkans to the ERIC loan class introduced and advocated here is reasonable.

### 4.3.2 Numerals

Numerals constitute a particularly telling area of study in language contact, as numeral borrowing may well be limited in contact situations. As Matras 2009: 201 puts it "Given that quantifying objects is considered a very basic human cognitive ability, it might seem surprising that many languages do, in fact, borrow numerals." And, there is the evidence of second language learners having difficulties with numerals, in counting in general but especially in the context of learning or using arithmetical skills in a second language.[125] Moreover, "quantity," a notion that presumably takes in numerals, is one of the semantic fields covered in the Leipzig Loanword Typology project (Haspelmath & Tadmor 2009; see footnotes 86 and 87 above), and as a field it ranks relatively low – in the lower half of those surveyed – in terms of conduciveness to being borrowed. This status for numerals is reflected too in the fact that the low digits occur on the Swadesh list.

It is thus of some interest not only for the Balkan lexicon but for the study of language contact in general that there are both numeral loans as well as numeral word-formation patterns in the Balkans that show contact effects. Numeral loans in the Balkans are mainly concentrated on numerals higher than the low digits 'one' through 'five.' Thus it might seem that numeral loans in the Balkans do not indicate anything special regarding contact in the region. We would argue to the contrary, however, and claim that loans involving numerals are significant here in that they show the depth of the penetration of contact languages into the surrounding languages. That is, even though there are languages with relatively restricted numeral systems, where low digits might be all that can be judged as basic, the Indo-European Balkan languages all have numerals up to 'ten' that are unanalyzable units, and thus on morphological grounds constitute "basic" elements of vocabulary; and even higher numerals, while built in a compositional way (except for StAlb and Aro '20'), nonetheless contain unanalyzable elements and have idiosyncrasies of internal ordering that give them a basic character. Thus, their involvement in transfer across Balkan languages is consistent with the intimate nature of much of what is surveyed here.

Numerals therefore represent a coherent lexical domain in which ERIC-type loanword contact effects can be discerned. We survey here the relevant evidence, consisting of some localized effects between two languages and a somewhat more widespread one that has been much discussed in the literature.

125 Greene 1992: 500, e.g., notes that "it is well known that it is very difficult for an adult to change the linguistic code in which he was first taught to manipulate numbers"; see also Galligan 2004.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.2.1 Localized Numeral Borrowing

There are several different pairs of languages in the Balkans involved in localized borrowing of numerals, as outlined in the following sections.

#### 4.3.2.1.1 Romani and Greek

In all Romani dialects, for instance, the numerals 'seven,' 'eight,' 'nine,' and 'thirty' are borrowed from Greek εφτά, οχτώ, εννιά, τριάντα, respectively, giving, e.g., in Agía Varvára Romani (Messing 1988), *efta*, *oxto*, *inja*, *tranda* (but for 'thirty', note Dolenjski Rmi *trideset* [<Slavic], Welsh Rmi *trin deš* [native 'three ten'], ROMLEX). This is true not just for Balkan Romani but for almost all European varieties of Romani (except as just noted), due to what Matras 2002: 210 calls "a qualitatively unique" impact of "Greek . . . during the Early Romani phase," suggesting that the language entered Europe through the Balkans during the Byzantine period. Specifically about numerals, Matras 2009: 202 notes that "Romani tends to retain an inherited word for 'twenty' and for 'hundred,' but often has Greek words for the numerals in-between, though many dialects tend to replace these higher numerals through loans from their contemporary contact languages." Various Balkan and South Vlax Romani dialects show this pattern, but 'forty' and 'fifty' tend to have Greek forms even outside of Greece, e.g., Arli Romani in North Macedonia and Kosovo *saranda* 'forty,' *pinda* 'fifty' from Greek σαράντα, πενήντα (ROMLEX,[126] cf. Boretzky 2003: 51; Messing 1988). The use of *saranda* in Romani even occurs in North Russian (ROMLEX). Matras 2009: 211 also documents the borrowing of the Greek ordinal suffix -το- into Romani, e.g., *dujto* 'second,' from (native) *duj* 'two.'

#### 4.3.2.1.2 Turkish Numerals in Balkan Slavic, Romani, Albanian, Aromanian

Turkish numerals are used in various ways in some of the Balkan languages. For instance, in the Balkan Slavic dialects spoken in geographic Thrace and Macedonia, Slavic-speaking Muslims – and Christians, to a lesser extent (Kodov 1935) – use Turkish numerals to varying degrees in regions with significant Slavic-speaking Muslim populations. In Pomak villages in present-day Greece, Turkish numerals are used for 'five' and above, with on-going competition between the Slavic-derived and the Turkish-derived forms for 'ten,' *désit* and *on*, respectively (Theocharidēs 1996a: 53). Šiškov 1936: 11 reports that in Dovan-Hisar (Dugan Hisar, Grk Aisými), in the Dedeagach (Grk Alexandroúpoli) region, all the numerals were Turkish. A similar situation obtains in some of the Romani dialects of eastern Bulgaria, e.g., (Varna) Gadžikano, Varna (Xoraxane) Kalajdži, and Kaspičan, where all the numerals above three are Turkish (Elšík & Matras 2006:

126 On ROMLEX, see Halwachs et al. 2007.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

170; Gilliat-Smith 1944). Šiškov 1936: 11 also reports that in villages and towns with Christian Bulgarian (and Macedonian) speakers, such as Gjumjurdžina (Trk Gümülcine, Grk Komotini), Smoljan, Nevrokop (now Goce Delčev), and Drama (Grk Dráma), Turkish forms of '100' (*yüz*) and '1/2' (*yarım*) are in use, and Grannes et al. 2002: s.v. list Turkish *bir* 'one' as "colloquial" in Bulgarian. For Macedonian, Jašar-Nasteva 2001: 113 cites *safar* 'zero' < Trk *sıfır* 'idem,' although the Macedonian usually means 'nothing.' Other numerals occurring mostly in nineteenth-century folklore cited by Jašar-Nasteva *ibid*. are *elli* '50,' *on* '10,' *beš* '5' *onsekis* (StTrk *on sekiz*) '18,' *on iki bin* '12,000.'

Friedman's observation cited in §4.3 above is worth repeating here: "the only Macedonian traditional part of speech lacking Turkisms is the numeral, although there are Turkisms in numerical expressions, e.g., *čerek* 'quarter' [cf. also Alb *çerek*], and Turkish numerals embedded in other parts of speech, e.g., *bešlik* 'five-grosch silver coin.'" Regarding this last form, note also Grk (Cretan) μπεσλίκι 'idem' (Orfanos 2014: 274), Alb *beshlëk,* Aro *beshlîc,* BCMS *bešluk*, BSl *bešlik,* Rmn *beşlic (*Polenakovikj 2007: 87), and the now-obsolete BSl *ikilik* 'Turkish coin of two kuruş' (Grannes et al. 2002: s.v.; Jašar-Nasteva 2001: 114), as well as Aro and Rmn *iuzluc*, Blg *juzluk* '100 para' and Mac *juzluk* '100 denar note' from Trk *yüz* '100' (Polenakovikj 2007: 144; Jašar-Nasteva 2001: 113). Moreover, some specialized counting practices have Turkish numerals, such as in Balkan Slavic backgammon, dice- or card-playing terms, e.g., *birlik* (in Galičnik, *birlok*) 'ace,' Aro *birlic* 'idem' from Turkish *birlik* 'unity,' cf. also Albanian *birllëk*, Rmn *berlic* (Grannes et al. 2002: s.v.; Polenakovikj 2007: 90; Jašar-Nasteva 2001: 113).[127] Some derivatives of Turkish numeral forms occur in Cretan Greek (Orfanos 2014: 276–277), e.g., from *bir* 'one': μπιρίς 'self,' μπιρί (in μπιρί σου και μπιρί μου 'one for-you and one for-me'), and μπιρί μπάχι 'at once' from Trk *bir baş* 'at once' (lit., 'one head'), inter alia, and from *bin* 'thousand': μπιν κερατάς '(someone) cuckolded a thousand times over,' μπιν κατεργάρης 'a huge trickster.' Similarly, in older or dialectal usage, Bulgarian and Macedonian show lexicalized forms containing Turkish numerals, such as *seksen sekis* 'much, many' (from Trk *seksen sekiz* '88'), *doksan-dukus* 'much, many' or the related *oksan-dokus* 'too much, an excess' (from *doksan dokuz* '99'), *dokuzbablija* 'born out of wedlock' (lit., 'having nine fathers'), *dokuzda* 'angry' (lit., 'in nine,' probably from an expression like '[having fallen] into nine [bad moods]' or *dokuz dağı* '[taking on a load of] nine mountains'; BER I: s.v.), *birki* 'some; a number of' (from *bir iki* 'one two'), among many others (Grannes et al. 2002: s.vv.; Jašar-Nasteva 2001: 113). Colloquial Alb *birinxhi,* Aro *biriŋĝi*, BSl *birindži* 'first-rate, swell, top-notch' all come from the Turkish ordinal *birinci* 'first' (Polenavokij 2007: 90), and similarly Cretan Greek has μπιριντζής (Orfanos 2014: 277). Finally, Turkish *üç* 'three' is the source of Albanian *yç* 'a game played with three stones.'

127 The Albanian reflects the WRT backing of i > ɪ in final closed syllables.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

#### 4.3.2.1.3 A Modern Albanian Secondary Usage

In a secondary functional domain, namely giving numbers over the telephone, as discussed by Friedman 2010c, in modern Albanian usage in Albania (but not in Kosovo), some Italian numerals are used; *tetë* 'eight' is replaced by *otto*, and *pesë* 'five' is replaced by *cinque*. These replacements are motivated by an interest in clarity, since the telephone cuts off high and low frequencies (such that *pesë* and *tetë* risk confusion), much as *niner* is used in aviation-derived usage in English (Hock & Joseph 2019: 154). While this example does not involve contact between Balkan languages, it does involve a Balkan language and moreover attests further to the possibility of the borrowing of numerals under the right ecological conditions (even if not ERIC-style conditions in this case).

### 4.3.2.2 Teens as 'X-on-TEN'

As a final numeral-related parallel, we turn to one that has received considerable attention for more than a century. This is the convergence involving the formation of the numerals from eleven to nineteen, cited in Miklosich 1862, Sandfeld 1930, and virtually all of the subsequent Balkan linguistic handbooks.

The basic facts are that in Albanian, Balkan Romance, and Balkan Slavic, the "teens" are expressed as DIGIT-'on'-TEN, i.e., with a digit, followed by a form of the preposition for 'on,' followed by a form of the word for 'ten,' thus additively giving the 'teen' as 'DIGIT on-top-of (i.e., beyond) ten.' Examples for such a "locatival" pattern (as Reichenkron 1958 calls it) include the following, for 'eleven' and 'sixteen,' to take just two of the nine numerals so constructed (Table 4.5).[128]

Table 4.5 *'11' and '16' in Alb, BRo, BSl*

| | | |
|---|---|---|
| Alb | njëmbëdhjetë | (cf. *një* 'one,' *mbi* 'on,' *dhjetë* 'ten') |
| | gjashtëmbëdhjetë | (cf. *gjashtë* 'six,' *mbi* 'on,' *dhjetë* 'ten') |
| Aro | unãsprãdzatse | (cf. *un* 'one,' *-sprã-* 'on' (cf. *suprã* 'above'), *dzaţe* 'ten') |
| | sheasprãdzatse | (cf. *ş(e)áse* 'six,' *-sprã-* 'on' (cf. *suprã*), *dzaţe* 'ten')[129] |
| Rmn | unsprezece | (cf. *un* 'one,' *spre* 'on,' *zece* 'ten') |
| | şaisprezece | (also *şasesprezece*; cf. *şase* 'six,' *spre* 'on,' *zece* 'ten') |
| Megl | unsprăţi | (cf. *un* 'one,' *spră* 'on,' *zẹáţi* 'ten') |
| | şasprăţi | (cf. *şasi* 'six,' *spră* 'on,' *zẹáţi* 'ten')[130] |
| Blg | edinadeset | (cf. *edin* 'one,' *na* 'on,' *deset* 'ten') |
| | šestnadeset | (cf. *šest* 'six,' *na* 'on,' *deset* 'ten') |
| Mac | edinaeset | (cf. *edin* 'one,' *na* 'on,' *deset* 'ten') |
| | šesnaeset | (cf. *šest* 'six,' *na* 'on,' *deset* 'ten') |

128 In the actual pronunciation of these forms, and in some instances, as even reflected in the spelling, these are 'DIGIT-on-TEN' only from an etymological standpoint, in that the component parts have undergone various reductions typical in compounding. These subsequent reductions are irrelevant here.

129 Vrabie 2000: 78; 618 gives some variant forms that are more greatly reduced: *únsprăţi, únsprã, úsprãs; şásprã.*

130 These forms are from Atanasov 2002: 221–222; in the Tsărnareca dialect, these forms occur without the final *-tsi*. Capidan 1925b: 155 gives the teens without a final *-i*, and with *-spre-* internally.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

This type of teen formation is not found in the other Balkan languages, thus neither in Turkish, nor in Romani, nor in Greek (generally speaking, but see below on the Postclassical period): Turkish shows simple concatenation of 'ten,' *on*, with a digit, e.g., *onbir* '11' (cf. *bir* 'one'), *onaltı* '16' (cf. *altı* 'six'), whereas the pattern in Romani is TEN-'and'-DIGIT, e.g., *deš-u-jekh* '11,' and the formations found in present-day Greek are concatenated DIGIT-TEN (δέκα) for '11' and '12' (έντεκα/δώδεκα) and TEN-DIGIT for '13' through '19' (e.g., δεκαέξι for '16'). Thus this pattern has the appearance of a concentrated lexical Balkanism restricted to a subset of the languages, specifically Balkan Slavic, Balkan Romance, and Albanian.

We say "lexical" here even though the pattern in these forms lies on the borderline between morphology, syntax, and the lexicon; in some ways it resembles the VERB-'not'-VERB pattern discussed in §4.1. That is, these numerals, represented as DIGIT-'on'-TEN, constitute both a coherent class of lexical items, albeit a small one, consisting of semantically related forms, and a set of derived combinations (historically, compounds), each with a compositional semantics of existing words. As such, these numerals have a clear internal syntax, but also, as historical compounds, they are lexical items. Given the enriched view of the lexicon advocated in §4.1, such a situation is not problematic, and it points to a need for flexibility in classifying phenomena (sometimes arbitrarily) regarding which domain of grammar they fall into.[131]

These facts have been widely discussed, in both handbooks and specialized studies, e.g., Reichenkron 1958. Their potential Balkanological interest stems from the comparison between these facts and the patterns for 'teen'-numeral formation in related languages outside the Balkans and/or at different stages in the development of the languages in the Balkans.

In particular, the formation of teen numerals via 'DIGIT-on-TEN' is pan-Slavic; compare the following, from East Slavic and West Slavic, in Table 4.6:[132]

Table 4.6 *'11' and '16' in selected non-Balkan Slavic languages*

| | | |
|---|---|---|
| Russian | odinnadtsat' (cf. *odin* 'one,' *na* 'on,' *des'at'* 'ten') | '11' |
| | šestnadtsat' (cf. *šest'* 'six,' *na* 'on,' *des'at'* 'ten') | '16' |
| Polish | jedenaście (cf. *jeden* 'one,' *na* 'on,' *dziesięć* 'ten') | '11' |
| | szesenaście (cf. *sześć* 'six,' *na* 'on,' *dziesięć* 'ten') | '16' |

131 Balkanists often label convergent or parallel features as "phonological Balkanisms," "syntactic Balkanisms," and so on.

132 See below for the Baltic facts and the interpretation they point to.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.7 *'11' and '16' in selected non-Balkan Romance languages*

| | |
|---|---|
| French | onze '11,' seize '16' |
| Spanish | once '11,' dieciséis '16' |

And, this pattern occurs in Balkan Romance but not the rest of Romance, as in Table 4.7, which are simply 'one-ten,' 'six-ten,' 'one-ten,' and 'ten-six' respectively, reflecting Latin numerals somewhat faithfully, e.g., *undecim* '11,' *se(x)decim* '16,' though with the order of elements reversed in the latter Spanish form. What these mean is that the Balkan Romance numerals represent an innovation away from the Latin situation.[133]

Greek too shows innovation in some of the teens, away from Ancient Greek, but, significantly, not in direction of other Balkan languages. For '13' through '19' in Ancient Greek, the pattern was 'DIGIT-and-TEN', e.g., τρεῖς καὶ δέκα for '13' (καί = 'and'). The modern asyndetic concatenation, e.g., δεκατρείς '13' ('ten-three'), is thus innovative within Greek. Moreover, it is the same as in Turkish, e.g., *onüç* 'idem,' suggesting that Greek was perhaps influenced by Turkish.

These facts have led most scholars to consider that what is found in Albanian and in Balkan Romance to be the result of Slavic influence on these languages, influence which did not extend to Greek or affect Romani after it entered the Balkans. There are however additional facts to consider that are relevant for evaluating this parallel, giving some reason to doubt its validity as a Balkanism per se.

First, as some have pointed out (e.g., Schaller 1975), Hungarian has a locatival pattern for the teens, with a locative case ending on 'ten' but with the digit following, e.g., *tizenegy* '11' (cf. *tíz* 'ten,' *-en* 'LOC', *egy* 'one'). Thus the order of elements differs from the Balkan pattern but the same elements are involved. The Hungarian pattern, however, is more extensive than the Balkan one; as Petrucci 1999: 133, footnote 30 states: "Hungarian also uses the locative pattern for '21' through '29': *husz-on-egy* '21,' *husz-on-ketto* '22,' etc. (Reichenkron 1958:162)," where *-on* is the back-harmonic allomorph of *-en*. Interestingly, the language that Hungarian replaced in its region was Slavic, so that this Slavic-type construction in Hungarian could well be a substratum effect in Hungarian (see also Chapter 6, footnote 5), as Slavic speakers shifted to the new language; the extension of the pattern to the twenties is understandable in language-internal terms as an analogical spread.[134] In the end, though, Hungarian is irrelevant for assessing the Balkan situation, albeit possibly as another instance of the borrowing, or contact-related

133 Recall the point made in §3.4.2.2 that what makes Balkan convergences especially noteworthy is that they (typically) represent divergences away from the languages' respective earlier states.

134 The same extension to the twenties is found in Aromanian, though not in Romanian or Meglenoromanian.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

spread, of numeral formations; moreover, even if an indigenous Hungarian pattern, all it shows is that a language can have such a pattern independently, raising the question of independent origin for the Albanian and Balkan Romance situations.

Second, Greek shows some evidence of a Balkan-style locatival pattern before Slavic speakers entered the Balkans. Hinrichs 1999b: 440–441, for instance, mentions Greek examples from the fifth century CE like τῆς τρίτης ἐπὶ δέκα '(of) the third upon ten' for '(of) thirteen,' and says they are "*nach dem balkanischen Muster*" ('according to the Balkan pattern'). And, Mihăescu 1977 mentions the sporadic occurrence of 'DIGIT-on-TEN' numerals in Postclassical Greek, such as τρεῖς ἐπὶ δέκα 'three upon ten' for 'thirteen' from the fourth century and δύο ἐπὶ δέκα 'two upon ten' for 'twelve' from the fifth century. A conclusion to draw here is that this pattern can arise independently without contact, since the dating of these formations is too early for Slavic influence.[135]

Even so, these facts do not seriously alter the picture concerning the grouping of all of Slavic, Albanian, and Balkan Romance. However the final additional consideration does undermine the assessment of this feature as a shared contact-induced trait within these languages during the period when they were in contact in the Balkans. Hamp 1992b has pointed out that the words for 'twenty' in Balkan Slavic and Balkan Romance and for 'thirty' in Albanian show that the numeral 'ten' is treated as masculine in Slavic but feminine in Albanian and Balkan Romance. The following facts show that Slavic, Balkan Romance, and Albanian have gender in the numerals 'two' or 'three' (for Slavic, the question of neuter gender assignment is irrelevant here) (see Table 4.8).

These forms figure in the multiples of 'ten' and show gendered forms, thus revealing a key difference in detail between the Slavic formation and the Balkan Romance and Albanian formations (see Table 4.9).[136]

To fully appreciate the significance of this Albanian/Romanian gender mismatch vis-à-vis Slavic, the Baltic numeral facts become important. Baltic offers a mixed picture: Latvian has the Slavic-type formation (cf. *vienpadsmit* '11,' *sešpadsmit* '16'),

Table 4.8 *Gendered numerals in Balkan languages*

| | | | |
|---|---|---|---|
| Slavic gendered numeral: | dva (M) | dve (F) | 'two' |
| Romanian gendered numeral: | doi (M) | două (F) | 'two' |
| Albanian gendered numerals: | tre (M) | tri (F) | 'three' |
| | (dy [M] | dȳ [F] | 'two') |

135 One could suppose some substrate influence here, and it would be interesting to know whether this and other examples were perhaps created by nonnative speakers of Greek. Mihaescu's sporadic Postclassical Greek examples go beyond the teens, e.g., ἑπτὰ ἐπὶ πεντήκοντα 'seven upon fifty' for '57' (fourth century).

136 The gendered forms of 'two' in Albanian are irrelevant here since Albanian has an apparent vigesimal basis for 'twenty,' *njëzet*, literally 'one-twenty' (and StAlb *dyzetë*, 'two-twenties' for 'forty'). We can also note here that OCS *desętь* 'ten' was sometimes treated as feminine. However, in 'twenty' it is always masculine.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.9 *Compound numerals with 'ten' in Balkan languages*

| | |
|---|---|
| OCS | dъvadesęti 'twenty' (lit., 'two tens,' with M *dъva*, thus M 'ten') |
| Romanian | douăzeci 'twenty' (lit., 'two tens,' with F *două*, thus F 'ten') |
| Albanian | tridhjetë 'thirty' (lit., 'three tens,' with F *tri*, thus F 'ten') |

whereas Lithuanian does not (cf. *vienuolika* '11,' *šešiolika* '16,' with an element –*lika*, from **leikw*- 'leave,' not a form of '10' (*dešimt*), generalized from the pattern seen in Germanic with '11' and '12' ('one-left (over),' 'two-left (over),' respectively, as if counting on one's fingers but working with a base-twelve system). Unfortunately, no teens are attested for Old Prussian. Thus Lithuanian sides with Germanic while Latvian sides with Slavic, pairings that make sense in terms of geography, assuming the present-day geographical relationship matches an earlier one, even if not in just the same locale.

Following Hamp 1992b, the interpretation of all of these facts for the Balkans is that Albanian only superficially has the Slavic (-Latvian) pattern, because it also has a different gender for 'ten' (although OCS '10' can also show feminine agreement, albeit not in the ordinal numeral). Hamp proposes that there was a period in which the variety of Indo-European which was to become Albanian (Albanoid) was part of a northwest Indo-European grouping in which Germanic and Balto-Slavic and Albanoid were in contact. Albanoid, along with Latvian, and Early Common Slavic, got the DIGIT-'on'-TEN pattern (presumably as an innovation in one group that diffused into the others) at this time, but altered it somewhat when it moved down into the Balkans and encountered the variety of Latin to which some speakers shifted, yielding Balkan Romance. In this way, Hamp accounts for the similarities between Albanian and Slavic (and Latvian), and the differences between Latvian and Lithuanian, while still allowing for the specific form of the Albanian-Balkan Romance parallel to emerge. And, it is supported in part by other features that link Albanian at a deep level with Balto-Slavic, especially the Winter's Law lengthening of vowels before the original voiced plain stops (see §1.2.3.1).

The occurrence of the DIGIT-'on'-TEN locatival pattern for the teen numerals in Balkan Slavic therefore has a different history from the pattern in Albanian and especially Balkan Romance. There is convergence, but it dates from a pre-Balkan period, and moreover, there is an important divergence to consider as well. More specifically, and more importantly for the evaluation of this parallel, this pattern cannot be a Slavic one that has been imposed on other languages of Balkans after their migration thereto, and in that sense it is not a (true) Balkanism.

### 4.3.3 Loans with Grammatical Value

Elements that serve a grammatical function, whether words or affixes, are typically part of tightly knit combinations that are not easily parsed in natural second language acquisition. Such function words are typically unaccented, adding to their being part

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

of the background of a phrase or sentence and not part of the outstanding elements (nouns and verbs). As to what can be called the glue of these latter syntactic units, holding content words together and showing how they relate to one another, they are among the items generally considered to be less easily acquired in second language acquisition and thus less easily borrowed in contact situations. In the Thomason & Kaufman 1988: 74–76 "borrowing scale," for instance, the borrowing of the function words "conjunctions and various adverbial particles" requires scale point 2 (out of 5) "slightly more intense contact," and others, such as adpositions, require point 3 "more intense contact."

In this section, therefore, we examine grammatical lexemes that have diffused across language boundaries in the Balkans, a phenomenon briefly exemplified by Sandfeld 1930: 21 as "*mots dits grammaticaux*" ('words considered grammatical'). We focus here on the very forms themselves, and thus distinguish this contact effect from calquing, where equivalent native items are substituted into foreign phrasal or constructional "templates" as models.[137] That these forms have spread is consistent with the claim of sustained and intimate day-to-day contact among speakers of different languages in the region, with concomitant bilingualism, what we see as sprachbund-conducive conditions.

#### 4.3.3.1 Pronouns

Pronouns seem to occupy a special place in contact situations. While many instances are known of borrowing of indefinite pronouns and, less so, of interrogative pronouns (see Matras 2009: 198–199), the wholesale importation of personal pronouns across languages, while documented (see Thomason & Everett 2001 and Matras 2009: 203–208), seems a much rarer event. Matras attributes this to the function of pronominals and to the fact that many nominal forms are often pressed into service in pronominal functions, especially to indicate complex social relations (e.g., honorifics) rather than simple referentiality per se. All of the examples cited by Matras involve close and fairly intense contact, as with the Molise Romani borrowing of the 3PL pronoun *lor* from Italian, though in the case of Pirahã apparently borrowing personal pronouns from Tupi-Guarani, there was not necessarily any bilingualism, only the near absence of pronominal use by the Pirahã, "suggesting that borrowing may have served a distinct referential purpose" (Matras 2009: 204). Nonetheless, it is fair to say that pronouns generally rank rather low on scales of borrowability, and pronoun borrowing would not be expected in casual contact situations. If pronouns are to move across language boundaries at all, intense and sustained contact would appear to be a suitable precondition.[138]

137 See §4.3 for a discussion of calquing, and §3.2.1.7 and this chapter, passim, for some essentially lexical examples; grammatical examples are to be found throughout Chapter 6, though some are signaled in §3.2.1.7.

138 Matras 2009: 206–208 gives his own pre-conditions, involving accidental similarity of forms and transparent agglutinative models, though he also suggests an "internal motivation" within Romani involving a recurring move toward renewing the 3PL pronoun, with borrowing being one "option for the renewal process to go forward" (p. 207). This is seen, e.g., in Berovo Burgudži Rmi *onnar*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

There are various Balkan loan phenomena that center on pronominals. They are discussed in the sections that follow.

#### 4.3.3.1.1 General

There are some instances of the borrowing of pronominal-like elements in the Balkans. Most of the languages, for instance, have borrowed the Trk *hiç* 'nothing': Alb *hiç*, BRo *hici*, Mac *ič*, Blg *hič*, Jud *hiç*, Rmi *hič*, Ottoman-era Edirne Greek χιτš (Ronzevalle 1911: 457). In addition to meaning 'nothing' this word can also be used as a negative intensifier of the type 'not at all.' Nonetheless, the form is pronominal in origin and seems to have spread without much resistance, a fact which might be attributed to the higher degree of "nouniness" it shows compared with deictic or personal pronouns.

A somewhat clearer case of borrowing involving pronominals is the occurrence of the Turkish demonstrative *bu* 'this' in nineteenth-century texts in Macedonian and in the Greek of Ottoman Edirne (Ronzevalle 1911: 266). The Macedonian use is always in Turkish-centered discourse – e.g., *Stambol bu, lesno aren čoek ne može da se najdi* 'It's Istanbul, you can't find a good person easily' (Cepenkov 1972a: 154) – and the Edirne Greek usage is restricted to the expression μπου κιμ (cf. Trk *bu kim*) 'who (is) this?,' with the interrogative pronoun κιμ (Trk *kim*) 'who?' as well. Nonetheless, given the strong familiarity that Greeks and Macedonians had in those times with Turkish, we feel confident in speculating that these uses were parsable and recognizable to the non-Turkish users, perhaps indexing Turkish ways through this usage.[139] Moreover, κιμ has a few other uses in Edirne Greek, e.g., κιμ ο 'who is this?' (with demonstrative pronoun ο from Trk *o*), and some indefinite uses dialectally in Bulgarian (Grannes et al. 2002: s.v.), e.g., repeated *kimi . . . kimi* 'some . . . others,' and with native interrogative pronouns as in *kim koj* 'someone' (cf. Blg *koj* 'who?'), which may reflect in part Turkish uses, e.g., *kimi* 'some (of them),' *kimimiz* 'some of us.'

Turkish *her* 'every' was borrowed into Macedonian as *er* and used to form generalized pronouns with native material, e.g., *er koj* 'everyone,' *er što* 'everything,' etc. (Jašar-Nasteva 2001: 115).

Even clearer yet, and of a somewhat person-related nature, is the borrowing of the Greek indefinite pronoun καθένας 'each (one),' along with its feminine form καθεμιά, into Agía Varvára Romani (Igla 1996: s.v.).

A truly personal and thus more grammatical instance involves the first-person singular possessive in Aromanian. Aromanian has *-m* for 'my,' e.g., *bărbáte-m* 'my man,' as well as *-nji*, e.g., *inima-nji* 'my heart.' The *-nji*, according to Papahagi

'they,' where *on* could be either native or Turkish, but the *-nar* is definitely from the Turkish plural marker *-lAr* with assimilation.

139 Thus, we see this as rather like the use of French *moi* in English, where the first-person reference is well-known, even if one is not particularly conversant in French (cf. the joke that runs *You are so pretentious! Who, moi?*).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

1974: s.v., is from the Latin dative pronoun *mihi*, which could be used to express possession, presumably via an intermediate stage **mnihi* (or **$m^n$ihi*). As for *-m*, Papahagi 1974: s.v. takes it to be from the Greek possessive pronoun μου 'my,' phonetically [mu], which in the northern Greek dialects that Aromanian would be in contact with would be simply -μ ([m], due to the regular Northern Greek loss of unstressed high vowels (see §5.4.1.5).[140] Thus *-nji* is the inherited form, and *-m* would be a later borrowing.[141] There do not appear to be any other pronominal forms, possessive or otherwise, that were taken over from Greek, so one can wonder why first person would be privileged here,[142] but the hypothesis of a Greek source for *-m* fits the available evidence, even if isolated in Aromanian. One can note that the Molise Romani pronoun borrowing is restricted to just one "cell" of the person/number paradigm. Sepeči Rmi *Devlam* 'O my God!' (*Devel* 'god,' VOC *devla* + Trk 1SG *-m*), Sérres Rmi *sarimiz* 'all of us' (Rmi *sar* 'all,' Trk *-imiz* 'our'), and Skopje Arli Romani *Fatmam* 'my Fatma' (proper name + Trk 1SG *-m*) also show pronominal borrowing (see Cech & Heinschink 1999: 150–153; Sechidou 2011; Friedman 2013b). The same is true of the *-m* in nineteenth-century Macedonian folk poetry, e.g., *devojkom* 'my girl' (Koneski 2021: 335). We note here too the polite second singular, Rmn *dumniata*, Alb (Geg) *zotnia jote*, Aro *afindi*, Grk (Sarakatsan) η αφιντιά (with northern raising of the unstressed /e/), itself a Turkism (cf. *efendi*, discussed in footnote 274) (Skok 1927: 166); see §6.1.4.3 for more on politeness and number in the Balkans.

Another instance of pronominal borrowing that interacts with grammar is in the Turkish of the Western Rhodopes (near Pazardžik and Smoljan), where the Bulgarian dative reflexive pronoun *si* and the masculine-neuter accustive *gu* (StBlg *go*) have been borrowed in the local Turkish dialect as in these examples (Mollov & Mollova 1966: 124–125, using their orthography): *čăkarăm si* 'I'm leaving,' *ben si giderim* 'I'm going,' *ajnada si bak* 'look at yourself in the mirror' (mirror.LOC rfl.DAT look.IMPV), *jazdăm gu* 'I wrote to him,' *al si gu* 'take it for yourself.' The accusative pronoun can even co-occur with the native form as a kind of object reduplication: *ben onu vermišim gu* 'I have given it to him' (I him.ACC give.1SG.PRF him.ACC). See also §6.1.4.1.5.

In Judezmo, Turkish personal pronouns are incorporated into the expression *Sen favlar, ben entender* 'You [Trk] speak [Jud], I [Trk] understand [Jud].' While this could be taken to be simple codeswitching, in theories of codeswitching that define a so-called matrix language, and do so based on predication (cf. Matras 2012: 382), this is in fact a Judezmo utterance with Turkish insertions. Such an analysis is

140 A form [mu] does occur in Aromanian (Papahagi 1974: s.v.), from Greek, presumably showing influence of Standard Greek on the local dialect that influenced Aromanian.

141 Other accounts of 1SG *-m* are possible. For instance, there are *m*-forms of the first-person singular pronoun, especially in the possessive adjective/pronoun *a meu* (of whatever origin – since *a njeu* also occurs, the *m*-forms could be learnèdisms based on Latin or even from Romanian) so that *-m* could be an analogical replacement for enclitic *-n'i* based on these "strong" (or at least other) forms with *m-*.

142 We can speculate that first person would be especially frequent in conversation, so that Aromanians presumably would hear such forms often from Greeks in interacting with them.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

strengthened by the connotational meaning of the sentence, viz. 'Speak to me in Turkish and I will understand you, although I do not speak it' (Bunis 1999: 90). At the same time, this example raises the problem of the border between borrowing and codeswitching, especially regarding the possibility of one-word switches (touched on briefly at the end of §3.2.1.6).

#### 4.3.3.1.2 Indefinite Pronouns and Adverbs

Sandfeld 1930: 128 observes that Albanian and Balkan Romance show a similarity in the formation of nonspecific indefinite pronouns that suggests possible ancient contact, namely the use of Alb *-do* and BRo *-va*, both formants based on 3SG 'want,' e.g., Alb *kushdo*, Rmn *cineva* 'anybody,' Alb *kudo*, Rmn *undeva* 'anywhere,' Alb *kurdo*, Rmn *cîndva/cândva* 'anytime.'[143] Sandfeld also cites Alb *çëdo* 'anything/something,' but in modern Albanian, *çdo* (< *çë*+*do*) means 'any, every' and can be combined with a variety of words, e.g., *çdo gjë* 'everything,' cf. also *ndonjë gjë* 'something/anything' (*ndonjë* < *në* 'if' + *do* 'want.3SG' + *një* 'one,' cf. colloquial Alb *ndo* … *ndo* … 'either … or … ' We can also note Rmn *ceva* 'something, anything, etc.' Specific indefinites of the type Alb *dikush, diçka, diku, dikur, disa* 'somebody, something, somewhere, sometime, some [quantity]' employ *di* '[who] knows?' (Meyer 1891: s.v.). The usage is similar to one of the etymologies suggested for the Common Slavic indefinite prefix **nē*, OCS *ně-* 'some-' – which is quite productive in BSl – according to which the prefix comes from a contraction of *ne vě* 'not know,' although other etymologies have also been proposed (Vasmer 1986–1987: III s.v.). The use of 'know' to form indefinites might be connected to the pre-migration contact of Albanoid and (Balto-)Slavic discussed in §4.3.2.2.

Constructions of a similar type are attested in Latin, e.g., *quamvis* 'anyhow,' *quōlibet* 'anywhither,' where the former has an element from *volō* 'want,' and the latter utilizes a different, but semantically similar, verb (*libet* 'it is pleasing'). The generalization specifically of 'want' and its extension to other indefinites, however, seems to have taken place in the Balkans, since these constructions did not become productive in Romance outside the Balkans. As Sandfeld 1930: 116 notes, in some Aromanian dialects, the Albanian particle is simply borrowed and attached to native material, e.g., *itsido* 'anything' (*i* 'and/or' + *tsi* 'what' + *do*), which then provides a base for other words, e.g., *caretsido*, *iutsido*, *cãndutsido* 'anyone, anywhere, anytime.' Aromanian also has native Romance forms for some of these, e.g., *careva/caniva, cúni-vá* (DAT/GEN.SG, but also nom), *iuva, tsiva,* 'some/any-one, -where, -thing' as well as a number of other constructions (see, e.g., Papahagi 1974: passim; Vrabie 2000: 53).[144] Meglenoromanian has *tsiva* 'some/anything.'

143 English translations of these words sometimes involve *some-* or *every-*, depending on context.

144 A form *can* also occurs in Aromanian meaning 'no one,' which Papahagi 1974: 310 takes to be from ModGrk καν 'at all.' He also sees the *-n-* of *cúni-vá* (p. 408) as due to the influence of Grk κανείς (also κανένας) 'any-/no-one' (on which see below).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

In the case of Common Slavic, based on the evidence of OCS, it appears that plain interrogatives were used as indefinites (Huntley 1993: 145). This is still the case in Slovene, which is unique in modern Slavic in its preservation of the original situation, although bare interrogatives as indefinites occur as contextual variants elsewhere in Slavic. The rest of South Slavic has an old optative use of preposed *bilo* – originally the neuter resultative participle of 'be' – as a possibility, e.g., *bilo koj* 'whoever, etc.' The adjectival postposed root *god-* 'suitable' (cf. Latin *libet* in *quōlibet* above) is used here in Balkan Slavic in an old locative adverbial form *gode* but in BCMS the old accusative adverbial *god* also occurs. Balkan Slavic also has a postposed new subjunctive/optative construction of the type *(i) da e* '(and/even) DMS[145] be.3SG'. BCMS has a variety of indefinite pronominal constructions (Stevanović 1986: 301), including preposed unstressed *ma*, which is part of the standard but appears to be a specifically Montenegrin feature (cf. also Fielder 2008). Lekhitic and East Slavic use various modal particles based on 'be,' e.g., Pol *-bądź*, Belarusian *-nebudz'*, Russ *-nibud'*, Ukr *bud'*, all from the singular imperative of 'be' (cf. OCS *bǫdĭ*); Belarusian and Ukrainian both also have *aby-*, based on a conditional marker, and Polish shows *byle-* (cf. *bilo* cited above) and *lada-* (semantically similar to *god-*). Czech uses the quantifier *koli-* and Sorbian has *žkuli-*, while Slovak uses *hoc(i)-, vol'a-, -kol'vek, bar(s)-*; of these, *hoc(i)-, vol'a-* are both semantically and historically connected to 'want' verbs.

Turkish has a wide variety of strategies. Romani borrows from various contact languages, although sometimes employing native elements, e.g., *neko* 'someone' = Slv *ne* + Rmi *ko* 'who,' *diso* 'something' = Alb *di* + Rmi *so* 'what,' *dišta* 'idem' = Alb *di* + Srb *šta* (Cech et al. 2009: 12, 40, 166, et passim); note also some calques, e.g., Skopje Arli *neso, neko* on the model of Macedonian *nešto, nekoj.*

Greek goes its own way here, with κανείς, a negative polarity item 'no one' that also means 'anyone' in interrogative contexts. It derives from καὶ ἂν εἷς 'and if-ever one' (Thumb 1912: 96; see also the detailed discussion in Horrocks 2014: 67ff.), although Balkan Slavic *i da e* 'and if/let is' is close to a calque.

Given this situation, it would appear that the formation of indefinites in Albanian and Balkan Romance might be a shared innovation from the period of contact with Latin, while the Slavic developments in general, although they took place after the migration of Slavic to the Balkans, are basically independent. We can also note here the Balkan Turkism (of Arabic origin) *filan/filjan* (StTrk *filân*) meaning 'some X or other, so-and-so, etc.'

145 Throughout this book, we follow Friedman 1985a in using the term Dental Modal Subordinator (DMS) to refer to that modal particle that in all the Indo-European Balkan languages begins with a dental consonant (Alb *të*, BRo *să, si, s'*, BS *da,* Grk να, Rmi *te*), is always modal, and is usually a subordinator (except when functioning as an independent (insubordinate) optative marker or, at least in some frameworks, when participating in future constructions).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

#### 4.3.3.1.3 Negative Pronouns and Adverbs from Interrogatives

As Sandfeld 1930: 157 observes, the derivation of negative pronouns from interrogatives is a Common Slavic feature shared with Albanian and Balkan Romance, e.g., Mac/Blg *nikoj, nikoga[š], nikade/nikăde, nikak[o]*, Alb *askush, askund, askur [rë], assesi* 'nobody, nowhere, never, nohow.' Sandfeld 1930: 157 cites the Banat Romanian GEN/DAT *nicicui* 'nobody,' and Romanian also has *nici unde* 'nowhere,' *nicicînd* 'never,' *nicidecum*/*nici cum* 'nohow.' We can also mention Meglenoromanian *nitiscari, nitsicăn, niccum* 'nobody, never, nohow.' Although Aromanian has *nitsi* (= Rmn *nici* < Lat *neque* 'not'), it uses sentence negation with indefinites to express negative pronouns. Note also the Aromanian dialect of Aminciu (Grk Métsovo), which has the form *cantsiva* 'nobody,' where the element *can-* is based on Grk κανείς (see §4.3.3.1.2, and footnote 144). Balkan Romani dialects borrow Slavic *ni* as in *niko* 'nobody,' *niso* 'nothing,' *nijekh* 'no one' (cf. *ko* 'who,' *so* 'what,' *jekh* 'one'), although Slavic *nikoj* is sometimes borrowed wholesale. The pattern also occurs in Baltic, e.g., Lith *nieko* 'nothing,' *nikas* 'nobody,' *niekada* 'never,' *nier* 'nowhere,' *niekaip* 'nohow.' Given the Balto-Slavic evidence on the one hand, and the far greater productivity of the pattern in Albanian than in Balkan Romance, on the other, we can speculate that, as with the 'on-ten' construction for teens (see §4.3.2.2), this might have been a Balto-Slavic/Albanoid northwest Indo-European areal feature that pre-dated the arrival of the respective ancestral languages in the Balkans.

### 4.3.3.2 Adpositions[146]

Adpositions are relational elements that pull pieces of an utterance together by marking how they relate to each other. They constitute a closed class of adverbials that mark specific grammatical functions, in some cases, syntactic arguments, but also, more usually, syntactic adjuncts. As such, they are part of the tightly knit combinations that serve grammatical purposes, thus fitting the profile of less easily borrowed items. And this resistance has been recognized.

Three adverbial notions often expressed by adpositions, especially in the Balkans, namely 'at,' 'in,' and 'with,' appear on the 207-word Swadesh list. Moreover, among the meanings ranked by Tadmor et al. 2010: 235 as showing little historical evidence of being borrowed in their forty-one-language sample, 'up' is ranked highest and 'behind' is twenty-fourth, and in their composite list of 100 vocabulary meanings that are "basic" by various measures (pp. 238–241), the adverbial notion 'in' occurs as number ninety-seven. All of this testimony, taken together, is consistent with the intuition that adpositions, especially those expressing local relations, are less likely targets for borrowing.

Nonetheless, as far as second language acquisition is concerned, Matras 2009: 29 considers "the choice of prepositions modifying objects" to be among the

146 See §4.3.10.2 on contact-induced shifts in prepositional semantics.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

"vulnerable categories" in a bilingual's codeswitching, and, perhaps relatedly, examples of the borrowing of prepositions do occur, including in the Balkans. Matras 2009: 200 claims this may be so especially for "expressions of more peripheral and more complex local relations," and in his listing of examples from various parts of the world, notes a few Balkan cases: Greek χωρίς 'without,' εκτός 'except for,' Romanian *în loc de* 'instead,' and various Slavic prepositions such as *bez* 'without' or *vmesto* (as in Bulgarian) or *namesto* (as in Macedonian) 'instead of' all occur in, *mutatis mutandis*, Romani.

To those examples, other Balkan cases can be added, e.g., colloquial Macedonian and older Bulgarian use of the Greek distributive preposition κατά 'x by x' with temporal expressions meaning 'every,' e.g., Mac *katadneven* 'daily,' *katagodišen* 'yearly,' *katautro* 'every morning,' etc. (cf. Sandfeld 1930: 21–22 and Gerov 1895–1908: s.v.).[147] There is also Alb *anámesa*, Aro *anámisa* '(in) between' from Greek ανάμεσα 'idem,'[148] and Alb *andis* 'instead' from earlier or dialectal Greek αντίς (now more usually αντί). Moreover, Sasse 1991: 320ff. gives instances of Greek prepositions in Arvanitika, including ανάμεσα and αντίς, as well as εκτός 'except,' εναντίον 'opposite,' μέχρι 'until,' and μεταξύ 'between.'

And these examples can be multiplied, with some interesting syntactic effects. Bulgarian has *gibi* 'like' from Turkish, as did Ottoman-era Edirne Greek (Ronzevalle 1911: 89). Macedonian and Bulgarian both have *karši* 'opposite' from Trk *karşı*, as does Albanian (*karshi*) and Aromanian (*carshí*), and so did Edirne Greek during Ottoman times (Ronzevalle 1911: 411). For the most part, the Turkish postpositions have become prepositions – as expected since the languages are primarily prepositional – so that the shift is in keeping with the general typological cut of each language. Still, there are exceptions: in Edirne Greek, the postpositional use of *gibi* is documented (e.g., in Ronzevalle 1911:89), and in Aromanian (Papahagi 1974:371), *carshí* is postposed (*doĭ oamenĭ carshí* 'opposite two men'). Moreover, *mene karši* 'opposite my place' occurs in Macedonian as a marked word order with the focus on *mene* (VAF field notes 2017). The final *–i* in all these forms for 'opposite' may well reflect a direct borrowing from a local (Balkan) Turkish form *karşi* (as opposed to being an adaptation of a form like standard Turkish *karşı*) since Aromanian copied the syntax of the Turkish and Edirne Greek seems in general not to have nativized the Turkish words it borrowed; nonetheless, in principle, in a given language, the word could have entered from a nativizing Balkan intermediary. A number of Romani dialects, e.g., Kaspičan, Sliven Nange, Kalburdži (RMS), and Futadži (Ivanov 2000), borrow a few Turkish postpositions as postpositions, and sometimes even extend this to native forms, e.g., Kaspičan *xəzmečestar sora* 'after work' (lit., 'work.ABL after'), which borrows the Turkish postposition *sonra* 'after' and calques the ablative governance of Turkish, but also, e.g., Sliven Nange *shtar zisendar palal* 'after four days' (lit.,

147 The more usual form in Greek now for this usage, κάθε (ultimately related to κατά), was also recorded by (Sandfeld 1930: 22) for Albanian and Aromanian.

148 The Aromanian is from a northern Greek source, as the vocalism *-i-*, with northern raising of unstressed *-e-*, shows.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'four days.ABL after'), where *palal* 'after' is native and the ablative on 'days' calques the Turkish case syntax. Gilliat-Smith (1915/1916: 87) records the use of the Turkish postposition *beri* 'since' as a postposition in various Romani dialects of Bulgaria, sometimes borrowed with the Turkish ablative case suffix that *beri* governs in Turkish, e.g., Kalajdži Rmi *račjardan beri* 'since the night began' (cf. Rmi *rat* 'night').

Matras 2009: 200 gives instances from Domari of the borrowing of "core prepositions" (from Arabic), and there may be such a case in the Balkans, involving 'with' in Greek and Albanian. The languages have identical forms, *me* in each, but the direction of influence, if any, is not clear. The relationship, if there is one, is complex, and so a more detailed treatment here demonstrates the difficulties in identifying historical relationships for some words.

First, the details of how με developed in Greek are not entirely clear. It is first attested in Medieval Greek (Hatzidakis 1905: 1.474; Bortone 2010: 221). The received wisdom (so Andriotis 1983: s.v., and all lexical compendia, e.g., Dangitsis 1978: s.v.; LKN: s.v.; Babiniotis 1998: s.v.; and Charalambakis 2014: s.v.), following Hatzdakis *op. cit*., is that it derives from Ancient Greek μετά by haplology and resegmentation operating on μετά with a neuter accusative plural nominal, either the demonstrative ταῦτα or a noun phrase introduced by the definite article τά; that is, μετὰ ταῦτα/μετὰ τά X . . . became μετ΄ Ø X . . . (or μεØ τα X . . .) and was then analyzed as a new form με with τα as the article.[149] While possible, this account looks ad hoc and is not entirely convincing.[150] Complicating the issue is the occurrence of an adverbial element με- in Ancient Greek, in composite forms like μέχρι 'until, up to' (ModGrk μέχρι) or μέσφα 'till.' This με- is surely an inherited element from Proto-Indo-European, as there are cognates elsewhere that parallel these Greek formations, e.g., Armenian *merj* 'close, near,' Gothic *miþ* 'with.' Conceivably, then, Modern Greek με could be an archaism, despite its late attestation, a relic of an independent use of με- not directly attested in Ancient Greek, or it could reflect the reanalysis of affixal adverbial με- to independent word status.[151]

149 AGrk μετά with the accusative means 'after,' whereas various senses of 'with' occur with the genitive or dative; however, in later Greek (as in Balkan Slavic and Balkan Romance), most prepositions came to take accusative only, so that the comitative (etc.) meanings of genitive/dative use could well have been associated with accusative syntax for some period of time when this putative haplology and resegmentation would have occurred. In Modern Greek one does find μετά with accusative meaning 'after,' and with genitive meaning 'with,' but these are borrowings from the consciously archaizing high-style Katharevousa into colloquial usage.

150 For instance, one can wonder how frequent neuter plural definites were with μετά and thus whether there would have been enough occurrences to induce such a reanalysis and to make it stay in the language.

151 Some linguists – especially advocates of grammaticalization who take a strong position against the possibility of movement from a bound element to a free element (the so-called "unidirectionality" issue – see Haspelmath 2004; Ziegeler 2004) – would say that such upgrading is impossible or next-to impossible. However there do seem to be well-documented cases of upgrading, including the movement of the prefix ξανα- 'again' to free adverbial word status in Greek (Dosuna 1997), and even Haspelmath 2004 accepts some such developments as legitimate counterexamples. Thus we do not consider it an unreasonable possibility here with με-.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

On the Albanian side, the occurrence of *me* could simply represent the Indo-European element directly, with expanded usage: *me*, presumably the same one as at issue here, serves to mark the Geg infinitive, occurring with a participle (e.g., *me punue* 'to work'), and a preverb usage can be detected in the verb *marr* 'take,' past *mora* 'took,' analyzable as **me* with the root found in AGrk ἄρνυμαι 'win, gain,' Arme *aṙnum* 'take.'[152] Some etymological sources on Albanian are simply silent on *me* (B. Demiraj 1997); Meyer 1891: s.v. says that it is a borrowing from Greek while Çabej 2014: s.v. summarizes the various arguments and opinions, siding with Jokl 1940: 128–129.

However, there are uncertainties on the Greek side, and there are disagreements between Greek and Albanian as to the functions of the respective *me*'s. Greek με, for instance, is not used as a preverb, though adpositions and preverbs do correlate across Indo-European generally (cf. Greek κατά 'down,' πρός 'toward,' and even μετά, among others), and no use of με in Greek parallels the Geg use as an infinitival marker. Moreover, Albanian *me* is not used in adverbial composites of the type of Greek μέχρι.[153]

These functional mismatches might point to independent origins for the respective *me*'s, perhaps independent inheritance from Proto-Indo-European. If a borrowing, despite Meyer's strong opinion, the directionality is not clear, and it could just as easily be that Greek borrowed from Albanian as vice versa. The fact that the prepositional use of με is not found in Ancient Greek is suggestive that this particular use represents a borrowing, especially since colloquial Medieval Greek was in contact with Albanian (Arvanitika) all over the Hellenic peninsula and the islands.

The semantic matching of the two *me*'s as prepositions is striking, in that each can mark means, accompaniment, and circumstance (and see also §4.3.10.2). Therefore the specifically prepositional uses may be borrowed, or at least modeled on one another, even if other functions do not match up directly. In the end, a definitive assessment of the history of this convergence cannot go beyond the speculations here, and we are left with a parallelism between Greek and Albanian that may or may not be contact-related.

The matching between Greek and Albanian in the semantics of με/*me* is instructive, as there is convergence all across the Balkans in preposition use and in the range of meanings for prepositions (cf. Asenova 2002: 97–104). With regard specifically to the semantics of Alb *me*, Grk με[τά], and BRo *cu* vis-à-vis BSl *sъ,* Asenova 2002: 101–102 observes that the Slavic preposition lost its ablative function quite early (it was taken over by Blg/Mac *ot/od*), thus bringing the BSl preposition in line with the other Balkan languages. The various functions that the

152 This etymology is due to Eric Hamp (p.c., see Hamp 2019 for details) and is somewhat simplified here; the *-rr-* in the present would reflect the presential nasal suffix seen in Greek and Armenian, with the *-r-* in the past reflecting the expected absence of the present stem marker in such a form.

153 Alb *megjithëse* 'although' and ModGrk μολονότι 'although' do not really count here since they are derivative from standard uses of their respective με/*me* forms and represent univerbations of original phrases headed by prepositional με/*me*. See §4.3.3.4 for more discussion of these.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

respective languages' prepositions currently have could have developed independently, but under the circumstances the parallels could only have been supported by language contact; see also §4.3.10.2 for more examples and discussion.

### 4.3.3.3 Negation

Negation clearly ranks among basic vocabulary material both in terms of its grammatical function and its discourse function. The meaning 'not' is on the Swadesh list, and, in Matras's 2009: 208 estimation, negation falls into the "essential and salient semantic relations that are likely to have some kind of structural manifestation in every language." He takes that as a reason why "not many examples of direct borrowing of word-form can be found" for this category. It is interesting, and telling, therefore, that Balkan languages show examples of the borrowing of negation.

There are several types of contact-induced developments involving negation-related items in the Balkans.[154] The most significant for the view of ERIC loans developed here, in that it is the most grammatical and thus the least expected under other than intense contact conditions, is the borrowing of the Greek negation marker μη.[155] This marker occurs in Greek with finite present tense verb forms, giving negative imperatives (prohibitives), e.g.: μη φύγεις 'don't (you.SG) leave!,' μη φύγετε 'don't (you.PL) leave!,'[156] and it is found in Balkan Slavic and Balkan Romance. Topolińska 1995a: 310 notes the occurrence of *mi* in the Macedonian of Lagadina (Grk Langadas) and sporadically elsewhere in Aegean Macedonian, and Stojkov 1968: 86 identifies it as characteristic of the Bulgarian of Strandža, in Thrace. Stankiewicz 1986a: 210 mentions *mi* as one of the "dialectal equivalents" (along with *nekaj*, *nemoj*, and *n'alaj*) in "various areas of Bulgaria and in some dialects of Macedonian and Serbo-Croatian" that can occur with an old truncated infinitive in a prohibitive expression. Papahagi 1974: 796 documents its occurrence in Aromanian, e.g., *mi γin'ĭ nculeá* 'Don't (you) come here.' The Modern Greek μη derives directly from Ancient Greek μή by regular sound changes ($\bar{e} > i$) and thus it has been part of Greek for millennia. There is nothing like it elsewhere in Slavic or Romance, so its occurrence in the Balkan varieties of these groups is best attributed to borrowing. While borrowing of negation in general may be unusual, the fact that it is prohibitive negation that is borrowed here finds motivation in the conversational and discourse basis of Balkan interaction that characterize ERIC loans, in that prohibitives can be expected to be particularly prevalent in conversational interactions.[157]

154 See also §4.3.3.1.1 regarding the borrowing of the Turkish negative pronoun *hiç* 'nothing.'

155 For discussion of parallels in the syntax of prohibitive markers in the Balkans, see §7.6.2.

156 These are perfective aspect forms; imperfective is possible with μη, e.g., μη μιλάς 'don't (2SG) keep talking.'

157 See also §4.3.4.3.2 and §7.6 for further contact-dependent convergences with other negation- and nonnegation-related uses of μη.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

There is also a prohibitive expression that is parallel across several of the languages, in some instances, just in regional dialects. As demonstrated by Papadamou 2019a: 796–797, drawing also on Papadamou & Papanastassiou 2013, there is a prohibitive structure in northern dialects of Greek (of Grammochoria in the Kastoria regional administrative unit) that involves a fossilized 3SG verb form φτάν' 'it is enough' (equivalent to StModGrk φτάνει 'it reaches; it is enough'), with a following verb, as in φτάν' κρέντς 'you spoke enough, do not speak' (2SG). She notes that this is "reminiscent of similar structures that carry the same function found in the local Slavic [i.e., Macedonian VAF/BDJ] dialects of the region, where, however, instead of φτάν,' the adverb *dosta* 'enough' is used, e.g., *dosta zborvi"* 'enough speaking, do not speak.' Moreover, she adds that "similar structures are also found in the Aromanian dialects of the region, where *duri/dure* (Papahagi 1974) is used as a marker of negation, which stems from the Turk. *dur* 'stop,'" e.g., *duri zburets* 'stop speaking, do not speak.' These parallels thus represent shared phraseology in the domain of negation that is somewhat grammatical in nature.

A second type of negation-related borrowing involves words that serve as a general statement of negation, an exclamatory utterance related to a discourse context that is equivalent in meaning to English *no* (and thus opposed to the grammatical *not*). There are two instances of such forms from Greek entering Southern Aromanian (Vrabie 2000: s.v. *no*): Greek όχι 'no,' giving Aromanian *ohi* (noted also much earlier by Récatas 1934[158]), competing with native *nu*, and Greek μπα, an interjection that means something like 'ah well' but also 'unh unh; no way,' functioning somewhat like όχι but more showing dismissiveness, giving Aromanian *ba* (also in Papahagi 1974: s.v.) in a similar meaning. Macedonian has the dismissive *ba*, while Romanian has *ba* as an exclamatory negator, as does Bulgarian ('of course not! Certainly not!' (Bojanova et al. 1998: s.v.)). Further, Agía Varvára Romani has *hayır* for 'no,' from Turkish (from the period before the speakers settled in a suburb of Athens). And, Manea 2013: 558 observes that *neam*, from (eastern) Bulgarian *njama*, occurs as "a Wallachian regionalism" in Romanian for 'not at all.' She further notes *ba* 'no, nay,' from Bulgarian *ba*, which, besides uses in combination with *nu* 'no' to express "the contradiction of an assertion," can also, "in non-standard contemporary Romanian [serve] as an archaic and colloquial variant of *nu* 'no' . . . especially with interrogative disjunctive clauses," e.g., *Ai fost la şcoală au* ***ba****?* 'have-you been to school or **not**?' (see §4.3.4.3.1 for more on *ba*, also found in Turkish and Judezmo).

Relevant here too is the noise-word, technically an ingressive voiceless dental affricate (alveolar click) – conventionally spelled *tsk* in English, *cq* in Albanian, *ck* in Macedonian and Bulgarian, τσουκ in Greek, *ţâţ* in Romanian, and *cık* (rarely *çık*) in Turkish, but all phonetically [ǀ] – which is the clucking noise that can accompany an upwards head-nod (downward in Balkan Slavic) for 'no' (on which see below). The ingressive velaric dental (dental click) negator occurs from India through the

158 As mentioned in §3.0, with reference to the mockery from mothers that the use of *ohi* by Aromanian children could occasion.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Middle East and northward into the Caucasus and northwestward into the Balkans and southern Italy and Sicily. Its northernmost extent in the Balkans appears to coincide with Ottoman boundaries, while in Italy it appears to coincide with the extent of Magna Graecia, and so is characteristic of southern Italy.

Another case like this is Turkish *yok*, which as a predicate means 'there is/are not' but is used also as a general emphatic negative exclamation meaning 'no' or 'absolutely not.' This has spread widely in the Balkans: note Balkan Romance *ioc*, Albanian and Balkan Slavic *jok*, Greek, as in Τουρκική η Κύπρος – γιοκ! 'Cyprus Turkish – No way!',[159] and also the former Serbo-Croatian, e.g., in a joke in which a Serb tells a Macedonian that Macedonians use lots of Turkisms, *a mi Srbi – jok!* 'but we Serbs, not at all!'.[160] The emphatic negator *jok* also occurs in Croatian sources (SANU 1973: s.v.; Matica Hrvatska 1967: s.v.). Further, *yok* may be involved in two additional instances of influence, as to use and, in one case, form as well.

The Albanian of Tetovo has uses for its inherited grammatical negator *nuk* 'not' in a way analogous to *yok*. The inherited grammatical negator *nuk* (pronounced [nawk] due to a regular diphthongization process in much of East Central Geg) has both functions of *yok*: it can be used to mean "there isn't any" and also as a one-word general negative utterance, roughly "no that isn't the case." While native *nuk ka,* (more frequently, *s'ka* in most of Albanian), with the existential use of the 3SG of 'have' (see §7.8.2.2.6), would mean 'there is none' and could in principle be reduced to simply *nuk*, the emphatic exclamatory use of *nuk* in Tetovo suggests Turkish influence. This is consistent with the status of Turkish as the urban Muslim home language in Tetovo prior to World War Two (Ellis 2003).

In addition, Joseph 2000c, 2001b, developing a suggestion made first by Landsman 1988–1989, argued that irregularities in the development of ModGrk όχι 'no!' from AGrk οὐχί 'not' can be explained by reference to influence from Trk *yok*. While όχι must derive in some way from οὐχί, the stress placement and the initial vocalism of όχι are unexpected, as is the functional shift from grammatical negation to a general exclamatory negative.[161] All three irregularities can be accounted for if *yok*, in its emphatic negative use, had an impact here, as the initial vowel and stress of όχι matches the stressed *-o-* of *yok*, and the functions match as well. Moreover, the chronology of the first appearances of όχι, in the sixteenth

159 This was observed as a newspaper headline in Greece in the fall of 1987. Though the Greek form is spelled γιοκ, its pronunciation matches the Turkish. With regard to the message, we can note that in Turkey, at that same time, one could purchase bumper stickers in Ankara saying *Kıbrıs türktür* 'Cyprus is Turkish.' As a more recent public example of γιοκ, one can note the roasted meat emporium founded in 2007 in Thessaloniki Γιόκ Μπαλίκ, literally "No Fish!", with the Turkish loanword μπαλίκ 'fish' (Turkish *balık*); see www.giokbalik.com/.

160 Albanian *jo* 'no,' though tantalizingly like Turkish *yok* (and note that *yo* occurs in Turkish as an emphatic negative utterance, as in *yo-yok*), is unlikely to be in any way related to it, since it occurs in Arbëresh, a dialect which shows almost no Turkish influence, since the speakers left the Balkans just as serious contact between Albanian and Turkish began (cf. Çabej 1996: s.v.; Joseph 2022b). On the other hand, it is worth noting that colloquially Albanian *jo* often ends with a glottal stop, whereas the affirmative *po* never does. See Hackstein 2020 for the suggestion that *jo* can be derived from a PIE **nēst* 'it is not the case; it is false.'

161 Ancient Greek ου should give Modern Greek ου ([u]) not ο ([o]), and the stress should have remained on the final syllable.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

century, accords well with this hypothesis of Turkish influence. Greek όχι would thus be a loan hybrid phonologically speaking, with the vocalism and stress of the Turkish form and the consonantism of the Greek (much like Tsakonian *ðon*, as discussed in §3.2.1.4). Again, given the exclamatory nature of όχι and *yok*, conversational interaction must have been the medium for such influence. Rijksbaron 2012, however, has addressed each of these points, finding evidence from the pre-Turkish era in Greek showing that each irregularity can be documented for Greek before Turkish influence was possible; he concludes that at the most, Turkish served to enhance the selection from among existing variants already present in Greek. (See also §4.3.3.1.1 on the Turkish negator *hiç* 'nothing, not at all.')

Finally, a third Balkan development involving negation is gestural in nature, thus paralinguistic, but still contact-related. Matras 2009: 196 is inclined to see gestural borrowing as part of discourse-related borrowing, drawing on the observation in Salmons 1990 about "the wholesale adoption of English discourse markers in Texas German as part of the overall convergence of communication patterns, including gestures," and we agree with this assessment.[162] Gestural borrowings in the Balkans are part of the conversationally based interactions associated with ERIC loans. Moreover, a gesture can only be borrowed if seen, so that gestural borrowing necessarily involves face-to-face interaction between speakers. The gesture in question here is the upwards nod of the head for 'no' – realized sometimes even as just the raising of the eyebrows – with an optional dental or alveolar click, as mentioned above. It is found in Greek, Balkan Romance, Albanian, Balkan Slavic (slightly modified in Bulgarian), and Turkish. The spread of this usage is thus clear, but the directionality is not. It occurs outside of the Balkans, e.g., in some Arabic speech communities (e.g., Lebanon), as well as Persian, and in India as well, suggesting it may have been imported into the Balkans through Turkish. At the same time, though, at least as far as Italy is concerned, it occurs in the south and in Sicily but not in the north, thus coinciding with the borders of ancient Magna Graecia, and therefore suggesting that it has been part of Greek for millennia (Morris et al. 1979). Nonetheless, whatever the source and direction of its spread, it clearly has diffused widely, and since it involves speakers interacting directly, face-to-face, it necessarily is tied to conversation, and is thus consistent with our ERIC loan rubric.[163]

162 Cf. also Joseph 2000c.

163 Some additional relevant data can be mentioned here. The Ancient Greek verb νεύω 'nod' occurs with the preverb ἀνα- 'up' in the meaning 'move head upwards' but also 'deny,' and with the preverb κατα 'down' in the meaning 'move head downwards' but also 'give assent.' Latin has a verb *nuō* that is cognate to νεύω that occurs with preverbs and gives meanings 'deny' (*abnuō*) and 'give assent' (*adnuō*). The preverbs have different basic meanings from the Greek ones; *ab* means 'away, off' while *ad* means 'to, into.' Still, it is interesting that one authoritative dictionary, the *Oxford Latin Dictionary* (*OLD*), gives for *abnuō* the specific meaning 'move the head (or eyes) upwards or away in token of refusal,' even in the absence of an 'up' meaning for the preverb. The Greek forms give greater concreteness to the directionality of the nod, and may thus indicate Greek as the source of the gesture, but the *OLD* definition suggests a possibly wider distribution for the gesture.

The possibility of contact involved in the negative gesture was mentioned as a relevant contact phenomenon in Joseph 2000c; see also Hauge 2002, where the dental click is noted as well. Gil 2013 shows that this use of the dental click is rather widespread, extending beyond the range of the upward head nod.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.3.4 Complementizers

Complementizers, or subordinating conjunctions, are part of what may be termed "clause-linking strategies," and they serve as markers in the discourse of crucial relations between clauses and, ultimately, utterances. They are lexical items but have grammatical and discourse functions. As such, they can be borrowed but their borrowing is tied to their role in discourse.

Matras 2009: 194, 196 gives numerous examples of the borrowing of certain discourse "connectors" (see below in §4.3.4.1 on these) which for him include complementizers, and suggests that "some of the most frequently borrowed subordinating conjunctions express concessive relations, causal relations, purpose, and conditionality," noting further that "factual complementisers appear to be more borrowing-prone than non-factual complementisers."

In the Balkans, the borrowing of such elements is well documented, with instances to be found of borrowed causal, factual, conditional, concessive, and some temporal subordinators. For instance, Agía Varvára Romani has temporal *molis* 'as soon as,' from Greek (Igla 1996: s.v.). Bulgarian has causal *zerem* 'because, since,' either from Gagauz (Grannes 1996: 144) or Turkish (standard Turkish *zira* 'because, since,' Grannes et al. 2002: s.v.) and Macedonian *zer, zere* 'idem' is also from Trk *zira* (Jašar-Nasteva 2001: 231), also BSl *čunki(m)* and dialectal Albanian *çynçi, çunçi, çimçi, qymqe* from Turkish *çünki* 'because' (Grannes et al. 2002: s.v; Jašar-Nasteva 2001: 36, 213, 230; Boretzky 1976: 38, 111). Both Bulgarian and Macedonian have *oti*, in the causal sense 'because, for that reason,' from earlier (Classical up into Byzantine and Medieval) Greek ὅτι, in the meaning 'for which/that reason.'[164] The entry of this form into Balkan Slavic is relatively early, predating the Ottoman period, as it occurs in the thirteenth-century Baniško gospel (Dogramadžieva & Rajkov 1981), so that it may have actually been a learnèd borrowing, found also in OCS (Sadnik & Aitzetmüller 1955: s.v.). Moreover, the causal meaning of ὅτι is available even today in Modern Greek (LKN: s.v. ότι2, Charalambakis 2014: s.v. ότι[2]), so that the chronology of the borrowing cannot be determined with any precision.

Another sense of ὅτι, meaning simply 'that' as a factual subordinate clause introducer, thus as a simple complementizer, was also borrowed into Bulgarian and Macedonian. This factual subordinator use is found in the Romani of Greece, too (Igla 1996: s.v.), and other varieties of Balkan Romani also borrow a local factual subordinator. In each case, as Matras 2009: 196 describes it, "the original Romani factual complementiser *kaj* is often replaced, in the respective dialects, by Greek *oti*, by Bulgarian *či* [< *če* with vowel reduction – VAF/BDJ], by Romanian-derived *ke*" and so on. He notes further that "the non-factual complementiser (inherited *te*) is virtually never replaced." This latter fact means that the Balkan distinction of factual versus nonfactual complementation (see §7.7.2.1.3.1), realized in Greek via ότι/πως vs. να, in Macedonian via *deka* vs. *da*, in Bulgarian via *če*

164 Formally, ὅτι is the neuter singular nominative/accusative of ὅστις 'anyone who; whosoever' (composed of the relative pronoun ὅς and the indefinite pronoun τις). It is regularly in Modern Greek now distinguished orthographically from ό,τι meaning 'that which,' which is made up etymologically of the same parts.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

vs. *da*, in Albanian via *se*/*që* vs. *të*, in Aromanian via *cã* vs. *si*, in Meglenoromanian *cã* vs. *s,'* and in Romanian via *că* vs. *să*, was carried over into and established, or perhaps maintained in, Romani through these borrowings.[165] We can also note here Gostivar WRT *se* 'that, because' borrowed from Albanian *se* 'idem' (Jašar-Nasteva 1970: 298). Here, Jašar-Nasteva makes the point that although the number of Albanian borrowings into Gostivar WRT is relatively small vis-à-vis Macedonian lexical items, such borrowings as exist tend to be function words.

In the area of concessives and conditionals, forms are borrowed that are based on the Greek word μακάρι, an old case form of Ancient Greek μάκαρ 'blessed' that in Postclassical Greek came to mean 'God willing,' and then took on grammatical use introducing wishes, in that way becoming complementizer-like. Judezmo of Istanbul (Varol Bornes 2008: 392) has *makaré* or *makari* with the imperfect subjunctive in the sense of 'if only,' e.g., *makaré fuera* 'if only this had been . . . .' Vlax Romani (Hancock 1995: 113) has *màkar kẹ* for 'although,' and *màkar te* for 'even if,' Macedonian, Bulgarian, and BCMS all have *makar* 'at least,' *makar što/če/da* for 'even though,' and *makar i da* for 'even if; although,' Aromanian has *macar(im)*[166] as an adverb meaning 'at least' but also as a connective, co-occurring with the subordinator *si*, meaning 'even if,' and Meglenoromanian has *măcar si* 'although' as well as *salde si* 'only if' (cf. Rmi *salde* 'only,' Mac *sal* 'only,' all from Turkish). For 'if' in Meglenoromanian, there is *acu* from Macedonian *ako,* and in Aromanian, one finds *ama că*, where *ama*, found all over the Balkans as 'but' (cf. §4.3.4.1), can here represent Greek άμα 'when, if.' In addition, Aromanian also has composite forms *s-easte că*/*s-fúre că* that are based on forms of the verb 'be,' with *fúre* deriving from the Latin perfect subjunctive of 'be,' *fuerit*. In that way, *s-fúre că* looks somewhat like Alb *në qofte se* 'if; in case that' (lit., 'in may.it.be (optative) that'), so that calquing – of uncertain direction however – is a possibility here. Meglenoromanian also borrows Mac *dali*, interrogative marker meaning 'if,' in the sense of 'whether.' Romani also has *dali* as well as *ako* 'if' from Macedonian and *eger* 'if' from Turkish (StTrk *eğer*).

Recognizing calquing and composite complementizers brings an Albanian–Greek complementizer parallel into focus, though most likely not one at the level of conversational usage. Both languages have composite forms for 'although' that are derived from prepositional 'with' plus 'all' plus the factual complementizer, as does Aromanian: Alb *megjithëse* (*me* + *gjithë* + *se*), ModGrk μολονότι (με + ολο- 'all' + οτι), Aro *cu tute cã*. The Greek, however, is a learnèdism, as the -ν- as a neuter singular ending is a *Katharevousa* feature, and the Albanian and Aromanian forms are likely calqued from that. Nonetheless, such parallels depend on a degree of awareness of the lexicon and grammar of the source language on the part of some recipient language speakers, and to that extent are indicative of one dimension of multilingualism in the Balkans.[167]

165 The equivocation here is due to the fact that although earlier Indic, as represented by Sanskrit and as suggested by Modern Indic, seems not to have made this distinction overtly through complementizer choice, there was a distinction between infinitival complementation and factual complementation with a finite verb.

166 The *-im* here is curious, but is presumably connected to the *–m* that occurs on the Turkish adverb *belkim*, a nonstandard variant of *belki* 'perhaps.' See also footnotes 181, 201, 204, 340.

167 See also §7.7.2.1.3.3, where further examples of such pan-Balkan composite complementizers are discussed, including the quite frequent 'for that' in purpose clauses.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.3.5 Interrogation

An interrogative marker serves a discourse function, but is also grammatical in that it is an indicator of a major sentence-type. As such, it would be expected to be somewhat resistant to borrowing, so it is significant that an overt marker for yes-no questions has been borrowed in the Balkans.[168] In particular, the Turkish postpositive *mI*, which is positioned in the string of postverbal elements before the personal endings, e.g., *türkçe biliyor mu sunuz* 'Do you know Turkish?' (= 'Turkish know.PROG Q 2PL') and shows vowel harmony with the verbal stem, was borrowed – in its rounded back harmonic form *mu* – into Edirne Greek during Ottoman times. Ronzevalle 1911: 451 describes it as "pleinement adoptée par les rouméliotes" ('fully adopted by the Roumeliotes') and gives the following examples:

(4.6) a. μπουρείς μου
can.2SG Q
'Can you (do it)?' (Standard Greek: μπορείς?)
b. θαρτ'ς μου
FUT.come.2SG Q
'Will you come?' (Standard Greek: θά 'ρθεις?)

This marker seems simply to be phrase- or utterance-final in the Greek, occurring after the personal endings (e.g., 2SG -(ει)ς), so that it has a somewhat different syntactic status from its Turkish source. Nonetheless, its grammatical marking function is carried over in the transfer.[169]

Various dialects of Balkan Romani show reflexes of borrowed Turkish *mI* as well as the Slavic yes-no question marker *li*, also borrowed, though in each case these interrogative markers are sometimes transformed into Romani evidential markers; see §6.2.5.3 for details and discussion. Slavic *li* is also borrowed as an interrogative marker into Aromanian (Cuvata 2009: s.v.). Turkish *mI* also occurs in other Balkan expressions (see §4.3.4.2.1).

### 4.3.3.6 Articles

Articles are among the least commonly borrowed elements cross-linguistically in terms of their form, as noted by Matras 2009: 216,[170] excluding cases where an article is incorporated in a lexical borrowing, as with Spanish *algodón* 'cotton' with the Arabic definite article *al-* carried along with the borrowed noun (*quṭun*). Interestingly, though, one case that Matras does cite is from the Balkans, involving

168 Meyer 1891: s.v. considers the Albanian yes-no question marker *a* to be a loan from Latin *an* but it is more likely a cognate of *an*, from PIE *$H_2en$ (cf. B. Demiraj 1997: s.v.).

169 Ronzevalle (p. 451) notes that other, more discourse-based functions, are carried over too: "*rend beaucoup de nuances concomitantes à l'interrogation : ironie, doute, défi, objuration*" ('[μου] offers many nuances that accompany interrogativity: irony, doubt, challenging, swearing (an oath)').

170 Other aspects of articles besides their specific form, such as the very occurrence of a definite or indefinite article or the positioning of the article (postposed or otherwise), can be a matter of language contact; see §6.1.2.2.1 on the much-discussed Balkan postposed definite article.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the Romani of Epirus in contact with Albanian. In that variety of Romani, an indefinite article with the form *njek* occurs, most likely "a blend of inherited Romani *(j)ek* and Albanian *një*" (p. 217).

It is worth mentioning here a parallel involving articles in the Balkans that is not quite as it seems. The Romani definite article (for more details on which see §6.1.2.2) has the form *o* in the masculine nominative and normally *i* in the feminine, which are quite strikingly identical to the Greek forms. Nonetheless, while some observers have intimated that contact with Greek may be involved in this parallel – Messing 1988: 18 says "It is probably not an accident that the Greek definite article for these two forms [Romani *o* and *i*] duplicates them" – the formal convergence is most likely merely a coincidence. Oblique forms of the article in some Central and Vlax Romani dialects preserve *l-* (< **t*) from the early Indic demonstrative source of the Romani article, and one can note that masculine nouns in Romani typically end in *-o* and feminine nouns in *-i,* so that those phones, on system-internal grounds, are associated with those respective genders (Sampson 1926: 247; Matras 2002: 96–98; Boretzky & Igla 2004: Map 47). At the same time, however, the deployment of the article in Romani, e.g., its use with proper names, does point to the possibility of early Greek influence (Boretzky 2000a).

There is one contact effect involving articles in the Balkans that is noteworthy. While the quite Balkanized Torlak dialects of eastern BCMS do have a (postposed) definite article, there is some retreat in the use of the article in the dialect of Niš and the Timok dialects in general and under normative pressure from standard Serbian, which does not have a definite article (Toma 1998).[171]

### 4.3.4 Discourse Elements

In the domain of ERIC vocabulary, we include those lexical items that serve as the "glue" of everyday conversational interactions between people, those holding discourse "chunks" together, much as function words hold syntactic chunks together (see §4.3.3). This covers a wide range of items, of varying functions. To some extent, all are interjectional in that they are not referential and do not mark actions or states or represent things; while some establish or signal overt connections between utterances, others seem to be fillers or hesitation markers. They include frequent discourse markers, which link utterances and often reveal a speaker's stance or attitude towards matters at hand, and indicators of an individual's status relative to other interlocutors that reflect solidarity and social distance more generally. In addition, they can modify the content of an utterance and can also serve a purely expressive purpose as elements that add "color" and "tone" to conversation.

It should be clear how these forms qualify as ERIC loans. As "discourse elements" they are necessarily tied to conversation and speaker interaction; indeed,

171 It is worth noting that the postposed definite article is emblematic of Balkan Slavic for BCMS speakers, who will say, for example, that Macedonian sounds like *tototo, tatata* (Friedman 2017a).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

as Brinton 1996: 33 remarks, discourse elements are "predominantly a feature of oral rather than of written discourse."[172] As such, they could not spread without such interaction, presumably on an intense and sustained basis, and since they generally show shared functionality,[173] a reasonable degree of bilingualism that would allow for the spread can be assumed. They are given a central role in the discussion of borrowing in Matras 1998, 2009 and in Hauge 2002, from which some of the material herein is drawn, and while we see them as important, for us they are one part of the larger conversationally based set of loans that we recognize. Nonetheless, although these discourse words are unified as markers that help to guide the discourse and move it along, they constitute a highly diverse set of elements; as a result, the classes proposed here are somewhat arbitrary and not completely discrete. In what follows, with a rough division into connective, modifying and expressive, and interjectional discourse elements, and bearing in mind that some elements can fit into more than one category,[174] we survey the diffusion of these discourse-related words in the Balkans.

#### 4.3.4.1 Connective Discourse Elements

A major aspect of conversation involves agreeing or disagreeing with an interlocutor, and particles marking these functions clearly connect speakers to previous utterances and indicate a speaker's stance. In §4.3.3.3, instances of the borrowing of such discourse-related (as opposed to grammatical) negators are presented, and here the affirmative side, particles meaning 'yes,' are added. Several instances are documented: Slavic *da* is 'yes' in Romanian, and according to Popnicola 1997, it occurs locally in the Aromanian of Bitola; Aromanian dialectally also shows *po*, from Albanian, and *malista* 'yes indeed,' from Greek (Vrabie 2000: s.v.).[175] Moreover, the Meglenoromanian spoken on the Greek side of the current border uses *ne* for 'yes' and *ohi* for 'no' (R. Atanasov 2016: 141, cf. Grk ναι, όχι). Albanian *po* also occurs in some dialects of Romani (e.g., Konopar Arli in Skopje). In earlier Bulgarian, though now obsolete, Turkish *evet* 'yes' is found, and dialectally the affirmatives *ăhă*, from dialectal Turkish *ıhı*, and *zar*, from Turkish *zira*, occur in the meaning 'yes; right' (Grannes et al. 2002: s.vv.).[176]

Quite widespread in the Balkans is a vocal gesture of affirmation with the form *e*. It is used in BSl to confirm something that someone else has said (i.e., not as an

172 This is not to say that it is not possible to study discourse elements in written language; in fact, Brinton 1996 devotes considerable space to discourse markers in Old English and Middle English, languages known now only through the medium of written texts. As for text-based Balkan languages, see Egea 1993, for instance, for some discussion of discourse markers in Medieval Greek.

173 Exceptions can be found in which diffused elements have different functions, but in such cases, the causes are identifiable, and interesting.

174 And could even go in a different section altogether; for instance, some material on negation in §4.3.3.3 is relevant here, as is the discussion on expressives in §4.3.4.2.2.

175 Greek μάλιστα was sufficiently well known that it served in bilingual jokes, e.g., in nineteenth-century Macedonia (Cepenkov 1972a:152; Friedman 1995b).

176 See also §4.3.4.2 on another use of *zar* (given there in the form *zer*).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

affirmative response to a discourse-new kind of yes/no question soliciting real information), in that sense functioning rather like English *right*. Albanian also has it, as does Greek, and in the Aromanian 'yes' (Cuvata 2006: s.v.; Papahagi 1974: s.v.). While it is a short utterance that undoubtedly belongs to the range of interjectional noises that humans can make universally, the functional match across these languages is striking.[177] What is not at all certain is the direction and source of diffusion, and it could well be that it was independently arrived at in each language but mutually reinforced through contact.

A connection of a different sort that also diffuses in the Balkans is the linking of chunks of discourse additively and adversatively. Matras 2009: 194 claims that such connectives show a "borrowability hierarchy based on contrast [of] but > or > and," so we take them in this order, though it is not always clear that this hierarchy is followed in the Balkans.

Certainly the connective with the greatest spread in the Balkans is 'but,' in that a subset of the forms *ama*/*ami*/*ma*/*mi*, each with various nuances of adversative value, occurring as a discourse marker and/or conjunction, is found in virtually all of the languages. Fielder 2008, 2009, 2010, 2015, 2019 has discussed this most thoroughly from the pan-Balkan angle, with other works that focus on the uses in particular languages. The distribution of the relevant forms is given in Table 4.10.[178]

The source is unclear and much disputed, as there are several plausible contact sources (e.g., Arabic into Turkish and then Turkish into other languages for *ama*, Italian into Greek and Albanian (and BCMS) for *ma*, Greek into other languages for

Table 4.10 *'but' in the Balkans*

| |
|---|
| *ama*, *ma*, *ami*, *mi* (as discourse marker and conjunction) |
| Aromanian |
| Greek |
| Bulgarian |
| Macedonian |
| Meglenoromanian |
| *ama*, *ma* only (as discourse marker and conjunction) |
| Albanian |
| Judezmo |
| Romani |
| Turkish |
| *ama*, *ma* (as discourse marker only) |
| Romanian |

177 Note, though, that Turkish *e*, is glossed 'enough!; well, all right; oh! (*surprise*)' (Redhouse 1968: s.v.).

178 Leaving out BCMS as per our focus; still, Bosnian is in the second group, while Croatian and Serbian are in the third group.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

*ami* (MGrk αμμή, etc.) as well as plausible internal sources in some cases (e.g., *ama* from Greek αμμή (from AGrk ἄν μή 'if + not') with final -α by analogy to αλλά 'but,' and μα by apocope from άμα), all of which make for the possibility of conflicting, and largely unprovable, claims.[179] For instance, άμα/*amma* occurs in Greek and in Turkish and given the extent of Greek influence in the vocabulary of Aromanian, the occurrence of *ama* there could be due to Greek; however, Turkish has also had a considerable impact on the Aromanian lexicon, so that *ama* could be a Turkism there, as indeed Vrabie 2000: 83 judges it.[180]

Besides the spread of the *(a)mV* word(s) for 'but,' other borrowing of adversative/contrastive connectives in the Balkans is attested. Meglenoromanian borrows *tucu* from Macedonian *tuku* 'but, rather.' Romani of Agía Varvára has borrowed αλλά 'but' from Greek (Matras 2009: 194) and *ala* occurs in Bulgarian and Macedonian sources though it is not much used in the present day, if at all (Fielder 2008: 116). A form *omos* 'however,' from Greek όμως 'however,' is reported dialectally for Macedonian by Budziszewska 1983. Also, several Turkish 'but'-like connectives are borrowed (some of which have other uses as well), see Table 4.11.

Finally, diffusion of 'but' is not restricted just to Balkan sources, if μα/*ma* in Greek and especially Albanian derives, as is quite plausible, from Italian *ma*.[181]

We can also note here the calqued adversative whose literal meaning is 'good but' and which has the semantics of 'however': Trk *iyi ama,* Rmi *šukar ama,* Alb *mirëpo,* BSl *[h]arno ama,* Aro *gine ama,* Megl *bun ama,* Grk καλά άμα. A similar convergent adversative is the use of 'and' with subjunctive (i.e., DMS + finite verb) to mean 'even if,' e.g., BSl *i da*, Alb *edhe të*, BRo *și s[ă],* Grk και να, as in Mac *i da dojdeš, fajde nema* 'even if you come, it's no use (lit., 'there is no profit')' (cf. also Sandfeld 1930: 108 regarding 'and' plus jussive in this meaning).

In the case of words for the disjunctive connective 'or,' there is one quite widely diffused form and other more localized borrowings. The widespread form is *ya* in

Table 4.11 *Borrowing of adversatives in the Balkans*

| | |
|---|---|
| ancak 'but, on the other hand, only': | Alb *anxhak* 'however,' Aro *anğeac* 'almost, finally,' Blg *andžak* 'precisely,' Mac *andžak* 'because' |
| illâ ve lâkin 'but on the other hand': | Alb *velakin*, Blg *illja veljakim*/*illjakim*, Mac *iljakim,* Aro *eleakim*/*ileakim*[182] |
| me(ğe)r 'but; however': | Blg *meger*/*mer* |

179 Fielder notes that there is an ideological (nationalistic) dimension to the positing of etymologies for these elements, with many linguists on all sides resisting a contact account.

180 Papahagi 1974: s.v. seems ambivalent here, mentioning both ("cf. Grk άμα 'sitôt que' < Trk *amma* . . . 'mais'").

181 Note too that Turkish *lâkin* is from Arabic, and further that Arabic is the source of 'but' in various non-Balkan languages, such as Swahili (*lakini*).

182 The final *-m* in the BSl and Aromanian forms here contrast with the *-n*# of the source (and in Albanian). Such interchange between final *-m* and *-n* occurs in other interjections and adverbials; see also footnotes 166, 201, 204, 340 on this. The expression is considered archaic today.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the expression *ya . . . ya* meaning 'either . . . or,' and it is found in Romani, Greek (για . . . για), Aromanian (*ia . . . ia*), Albanian, Balkan Slavic and Meglenoromanian (*ja . . . ja*), and Turkish (*ya . . . ya*). The locus of diffusion for these languages is surely Turkish,[183] as Matras 2009: 194 has it, commenting on Romani of Agía Varvára, though the presence of the word in Greek means that the Agía Varvára source in principle could be Greek as the local and "current contact language" rather than Turkish as "an older contact language." Other cases where the source is clear are the appearance of Macedonian *ili* 'or' and *a* 'or, whereas' in the Turkish spoken in North Macedonia and in Aromanian (Papahagi 1974: 675) as well as *ili . . . ili . . .* 'either . . . or . . . ' in Meglenoromanian, the occurrence in Albanian and in SDBR of *i* from Greek ή 'or' and *yohut* in Albanian from Turkish *yahut* 'or; otherwise' (Dell'Agata 1966), and the borrowing into Bulgarian of *ha . . . ha* and Macedonian *a . . . a* from Turkish (Grannes et al. 2002: s.v.; Jašar-Nasteva 2001: 229).[184] Similarly, Agía Varvára Romani (Igla 1996: 296) has the negative disjunctive connective *ne . . . ne* 'neither . . . nor,' from Turkish.[185]

The additive/conjunctive connective 'and' also yields examples of borrowing in the Balkans. Macedonian simplex 'and,' *i*, is borrowed into Aromanian,[186] and also into the Turkish spoken in North Macedonia, whence there is "reciprocity," in that the doubled *hem . . . hem* of Turkish, meaning 'both . . . and,' enters Macedonian, Aromanian, and Meglenoromanian (as *em . . . em . . .*), as well as Albanian, Bulgarian, Romanian, and Romani; single *[h]em* 'and, too, and yet,' occurs in all these languages was well as in the Greek of Ottoman-era Edirne (Ronzevalle 1911: 456). A borrowed form that is additively connective in that it keeps the discourse flowing is the Turkish *demek* 'that is to say, namely,' found throughout the Balkan languages, including Romani (e.g., Igla 1996: s.v.), Albanian (Boretzky 1976), Aromanian (Papahagi 1974: s.v. *demec*), Meglenoromanian, Macedonian, and colloquially in Bulgarian (Grannes et al. 2002: s.v.) and dialectally in Greek (LKNonline: s.v. ντεμέκ).[187]

The widespread occurrence of the *amV* word(s) for 'but' along with the wide distribution of *ya* for 'or' and the more restricted borrowing of 'and' would seem to suggest that Matras's hierarchy mentioned above is suitably instantiated in the Balkans. However, it is important to realize that in a language that attests two of these connectives, they need not have been borrowed in the sequence predicted by the hierarchy; in the absence of appropriate historical records, one has to be agnostic.

183 Turkish, though, has gotten it from Persian (Redhouse 1968: s.v.).

184 Presumably from *ha* in marking a speaker's stance on continuing activity (see §4.3.4.2.2).

185 This entered Turkish from Persian (Redhouse 1968: s.v.) and so could have entered Romani before contact with Turkish.

186 Although no source makes this explicit, presumably *i* 'and' was borrowed into northern Aromanian (with Macedonian as the main contact language) and *i* 'or' was borrowed into southern Aromanian (with Greek as the main contact language). Slavic *i* 'and' can also be encountered in Romani.

187 See below in §4.3.4.2.1 on a modifying use of *demek* in Albanian and Macedonian. Some Istanbul Judezmo speakers have *ne demek* 'What is that supposed to mean?!' as a fixed expression for intense disapproval (Varol Bornes 2008: 457).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

We turn now to two case studies which are sufficiently complex to require presentation in considerable detail.

#### 4.3.4.1.1 Detailed Case Study A: 'either . . . or'

The disjunctive conjunction meaning 'either . . . or . . . ' also has a specifically Balkan realization in the Albanian dialects of North Macedonia, which comes from Macedonian and may in turn be connected with Aromanian.

The Macedonian verbal *l*-form is descended from the Common Slavic resultative participle, which in Old Church Slavonic (*ceteris paribus,* the equivalent of Common Slavic) was used to form the perfect, pluperfect, conditional, and future perfect. In Macedonian, unlike Bulgarian, the *l*-form lost its ability to function attributively but remained in use for the perfect, pluperfect, and conditional. At some late stage in Common Slavic, and thus well before the rise of the opposition confirmative/non-confirmative in Balkan Slavic (see §6.2.5.1), what was the *l*-participle developed an optative usage in the third-person singular to replace the third singular imperative which, being homonymous with the second singular imperative, was lost. According to Vaillant 1966: 97, such usage is found in Czech as well as throughout South Slavic (cf. the common BCMS toast *živ[j]eli!* 'may [we] live'), and thus it must have arisen prior to their separation. For Polish, too, Topolińska 2008 points out uses of *bylo* that also look optative, as in example (4.7):

(4.7)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| a. | było | nie | było, | zrobimy | to | | | | (Pol) |
| | was | NEG | was | do.PFV.PRS.1PL | it | | | | |
| b. | kako | da | e, | kje | go | napravime | toa | | (Mac) |
| | how | DMS | is | FUT | it | do.PFV.PRS.1.PL | it | | |

'no matter what (Polish 'let it be/not be'), we will do it'

She compares this to uses of plain *bulo* protases in Ukrainian which can have an optative interpretation (Topolińska 2008: 172). Moreover, she notes that this usage occurs in eastern dialects, where the influence of Polish is unlikely, as in (4.8):

(4.8)

| | | | | | | |
|---|---|---|---|---|---|---|
| Buło | pryiti, | to | ja | skazała by . . . | | (Ukrainian) |
| was | come.INF | then | I | say.COND | | |

'If [someone] would come, I would say . . . '

In Macedonian, the old perfect using the *l*-form developed a chief contextual variant meaning of nonconfirmativity in opposition to the synthetic aorist and imperfect, which became markedly confirmative, i.e., denoting events for which the speaker is willing to vouch. The old perfect can thus be used to express surprise (or doubt, etc.) at a newly discovered state of affairs that existed before the moment of speech but that the speaker just discovered, e.g., *Toj bil tuka* normally means 'he was/has been here,' but it can also mean 'He is here (to my surprise)' and in this meaning it corresponds to Albanian *Ai qenka këtu* (see Friedman 1981, 1986a, 2005a for details).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Vaillant 1966: 97 attributes the optative uses of the *l*-participle to an elliptical optative composed of *da* plus the conditional (3SG *bi* plus *l*-participle), e.g., Macedonian *Dal ti Gospod dobro!*, literally 'May the Lord grant you [that which is] good!'. He also notes that Russian uses of the type *pošël* 'Let's go' have nothing to do with the South and West Slavic phenomenon under consideration here but are rather expressive uses of the past. (Cf. colloquial English *We're outta here*.) It thus seems to be the case that we are dealing with an old isogloss that spread from South to North to include West Slavic and even Ukrainian, but not Russian.

In Macedonian, the optative use of the *l*-form was reinterpreted as a perfect rather than an elliptical conditional and can thus occur in other persons with the auxiliary of the old perfect rather than the conditional marker, e.g., *Da ne sum te videl!*, literally 'May I not have seen you!,' i.e., 'I'd better not see you [around here].' In the course of subsequent centuries, the perfect meaning of the old present resultative perfect using the *l*-form in Macedonian came into competition with that paradigm's nonconfirmative meaning, which arose as a result of the development of marked confirmativity in the synthetic pasts (see §6.2.5.1 and Friedman 1986c for detailed discussion). In southwestern Macedonian, with the rise of a new resultative perfect using the auxiliary *ima* 'have' and the neuter verbal adjective, the old perfect using the present of 'be' plus the verbal *l*-form became restricted to nonconfirmative usage and, in the extreme southwest, disappeared almost entirely. To the north and east of the Ohrid-Struga region up to the river Vardar (and beyond, since World War Two), the old and new perfects have been in competition, and the old perfect using the verbal *l*-form is an unmarked past, but with a chief contextual variant meaning of nonconfirmativity (see Friedman 2014b:101–108 for detailed explanation, also §6.2.5.1). At the same time, with all these developments, a remnant of the old Late Common Slavic use of the *l*-participle as an optative (without, importantly, an auxiliary in all the languages where it occurs) developed in Macedonian and Bulgarian into a disjunction using the third-person singular neuter of 'be' as *bilo . . ., bilo . . .* (lit., 'let it be . . ., let it be . . . ') in the meaning 'whether . . ., or . . . ' (cf. archaic English *be he alive or be he dead . . .*).[188] In its meaning, this construction corresponds to the Albanian use of the 3SG present optative *qoftë . . ., qoftë . . .*. In modern Albanian, the optative is more or less limited to expressions such as *rrofsh!* 'thank you' (lit., 'may you live'), *me nder qofsh* 'you're welcome' (lit., 'may you be with honor'), and a variety of other formulae, blessings, and curses; these can use any verb in any person, so that even though quite restricted in function, the category is very much alive. This function, however, is very tightly connected to the desiderative function of the optative. As such, it rarely occurs outside this function, and when it does, e.g., in the expression *në qoftë se* 'if,' it can always be replaced by some other locution (*në, po, po të,* etc.).

188 Some speakers of BCMS accept the *bilo . . . bilo . . .* construction, and it is attested in literature, but many modern speakers today reject such usage.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

In the Albanian of North Macedonia (but not that of Kosovo, Montenegro, Albania, or Greece),[189] it appears that the combination of the general restriction of the Albanian optative to wishes combined with the surface similarity of the Macedonian optative use of the *l*-form to its nonconfirmative use, especially with the verb 'be,' has resulted in a calqued replacement of *qoftë* by *qenka* in the meaning of 'whether . . ., or . . . .' Thus, for example, an Albanian politician from Tetovo, talking with a colleague in Skopje about the importance of investment, made the point that nationality was irrelevant: *qenka shqiptar, qenka amerikan, qenka maqedonas* . . . '[it doesn't matter] whether it's (= let it be) an Albanian, an American, or a Macedonian . . . .' The Macedonian for *qenka* here would be *bilo*, while standard Albanian would use *qoftë* in this position (Friedman 2012b).

An Aromanian equivalent expression for 'whether . . . or . . . ' is *furecă* (*furică, furică, furcă, fucă*) as in *fure-că-i bărbat i fure-că-i mul'are* 'whether it be a man or a woman' (Capidan 1932: 511).[190] This corresponds to the Balkan Slavic *bilo . . . bilo . . .*.

From the point of view of Aromanian, *furecă* is an archaism, preserving the Common Balkan Romance (im)perfect conditional-optative of 3SG 'be' (Latin perfect subjunctive *fuerit*). From the point of view of Latin, however, it is an innovation on two counts. First, the Common Balkan Romance transformation of the imperfect and perfect subjunctive into a conditional-optative is an innovation, since Latin did not have a specific conditional paradigm, although the imperfect subjunctive was one of the tenses used to render conditional meanings (Rosetti et al. 1965: 184; Papahagi 1974: 67; Ivănescu 1980: 155, cited in Nevaci 2006: 143). Papahagi 1974: 67 makes the point that the merger of the perfect and imperfect subjunctives occurred in Common Balkan Romance. Moreover, the present and past synthetic conditionals have also merged for most or all verbs, with the new analytic conditional with *volo* 'want' replacing the synthetic past conditional, and, usually, the present conditional as well.[191] But precisely in the auxiliaries, the temporal opposition is preserved in form even if not necessarily in content, so that, formally, the 3SG present conditional of 'be' is *[s'] heare* and the (im)perfect is *(s[i]) fure*.[192]

The second innovation from the point of view of Latin is the use of this form of 'be' in a disjunctive alternative conjunction, where Latin had *sive . . . sive . . .* (rarely *seu . . . seu . . .*), which in turn comes from the locative of the demonstrative

189 This usage is limited to North Macedonia (Rexhep Ismajli, p.c.).

190 Vrabie 2000: 730 also notes *fureşi că*, and *ai că fure*. Capidan 1932: 509–510 also gives older *se-fure-că,* as well as *s-fure-că,* and, like Vrabie, notes that *sh(i)* 'and' can be inserted between *fure* and *că*. Saramandu 1984: 464 gives *s-fúri cî* and *s-eásti- cî*.

191 In modern Aromanian, the synthetic conditional appears to be rare or moribund. Thus, for example, it is not recorded in the Republic of North Macedonia (Gołąb 1984a; Markovikj 2007), and while Beis 2000: 334–336 records it for Aminciu (Grk Métsovo) on the western flank of the Pindus range, Bara et al. 2005: 200–201 do not find it in Turia (Grk Krania), northeast of Aminciu on the eastern flank of the Pindus. Weigand 1888: 95 also notes that the synthetic conditional is absent from Vlahoclisura (Grk Klisoúra), but records forms from Samarina (Grk Samarína), on the western flank of the Pindus. See also Nevaci 2006: 142–152.

192 Capidan 1932: 487, 509 treats all synthetic conditionals as present, and the past conditional is only analytic.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

**so-* plus the clitic conjunction *-ue* 'or.' We can also note here that the more common meaning of *(s)furecă*, etc. is simply 'if' (Romanian *dacă*; cf. Capidan 1932: 509; Saramandu 1984: 464), which meaning likewise had a very different form of expression in Latin. Thus far, the developments we have noted for Aromanian are suggestive, but only that. These developments occurred at a time when Romance, Slavic, and Albanian speakers were in contact with one another in the same place, and their verbal systems were all undergoing significant restructuring. As Gołąb 1976, 1984a, 1997 has shown, the influence of Aromanian on Macedonian, especially in the verbal system, was especially strong. At the same time, Romance-Albanian and Slavic-Albanian contacts are all well attested in the respective lexicons. Moreover, Scărlătoiu 1980 argues that Slavic-Romance contact occurred over a wide area, which means that innovations could also expand broadly. On the other hand, the Aromanian development is quite distinct from Romanian, which uses the Balkan Romance subjunctive of 'be' for the correlative alternative conjunction – *fie . . ., fie . . .* (cf. French *soit . . ., soit . . .*) – and has very different developments for 'if.' In all three languages, it is a past stem of 'be' that moves into this equivalent type of modal usage, while the conditional of the Balkan type (Gołąb 1964a), using *volo* plus imperfect marking, made significant inroads later into all three language systems, but did not completely eliminate the earlier constructions. We thus have a picture of complex accommodation and resistance.

The modern Albanian-Macedonian interaction casts light on the situation a millennium or so ago. It was a time of considerable change in the Romance, Albanian, and Slavic verbal systems as well as lexicons, and while the developments are not completely isomorphic, their parallels are striking. If Albanian was already beginning to develop its optative around the time that the dialects that became Common Balkan Romance lost contact with Latin, this might have given an impetus for the reinterpretation of the past subjunctives as a new distinct conditional-optative. This in turn might have influenced the optative development in the South Slavic perfect, which was early enough to spread north before the Magyar and German invasions cut off contact between what became South and West Slavic. Finally, it is worth noting that Greek does not have this type of correlative alternative conjunction and in general it is absent from these developments. This contributes to the idea that they are quite old, i.e., before Greek began to re-enter the hinterland from the coast after the Slavic migrations.

#### 4.3.4.1.2 Detailed Case Study B: An Expressive Connector

Finally, since this subsection treats connectives and §4.2.4.2 treats attitudinal expressives, it is appropriate to consider the Balkan particle *de* as it not only serves as a connector, but also expresses an attitude, usually a kind of emphatic; thus, it iconically provides a suitable connective in itself to the next section.

The relevant facts are that both BSl (including eastern Štokavian) and Turkish each has a native particle *de,* and in both the particle can be independent or enclitic. In both language groups the independent form can be derived from a verb of saying

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

and at the same time there is a homonymous particle of nonverbal origin. For BSl, the source of the 'say' particle is the Indo-European root **dhē-* 'do, put, etc.,' which gives OCS *děti* with an imperative *dej* that survives in the Bulgarian prohibitive *nedej* 'don't!' as well as in the archaic form of more recent *de* (BER I: 334). In BSl, the reportative meaning found, e.g., in Russ *de*, Ukr *di,* Pol *dzie* (Vasmer I: s.v.), does not occur. According to Skok 1971: s.v., Slavic *de* is of pronominal origin (IE **t-*), despite difficult historical phonology, although he also identifies clitic *de* as a Balkan Turkism. Skok notes that *de* can be used with imperatives, and in the plural imperative can even come between the verb and the plural marker: *dajde!* 'c'mon give! (2SG)' / *dajdete!* 'c'mon give! (2PL)'. He identifies this use of *de* with nonimperatives as typical of Kosovo for BCMS, e.g., *znamde!* 'Hey, I know, already.'

For Turkish, the Common Turkic root for the verb is **dij* 'say, etc.' For Turkish, *de!* (i.e., the vocalically invariant, bare, stressed verbal root) is still the imperative of 'say' as well as a freestanding expressive particle (Sevortjan 1980: 221–222; as a freestanding particle, Redhouse 1968: s.v. marks it as 'provincial' and glosses it 'Now then! Come on!'). Turkish also has an inherited enclitic particle *dV* with low vowel harmony, i.e., realized as *da* or *de* depending on the last vowel of the preceding item. This particle has the basic meaning of a coordinating conjunction 'and' but by extension is also emphatic with meanings like 'even,' 'even though,' etc. (Sevortjan 1980: 109–110). It is this second *de* that occurs in expressions such as *[h]em de* 'and also,' *ben de* 'me, too,' *bana da* 'me.DAT too.'

In Aromanian and Meglenoromanian, Albanian, and northern Greek, independent *de* has the exclamative meaning ('Hey!' or 'Well, now!) found in Turkish and Balkan Slavic (Skok 1971: s.v, cf. 1974: s.v.). Clitic *de* occurs chiefly with imperatives in Albanian and Greek, a usage that was noted above for Balkan Slavic and which also occurs in Balkan Romance and Turkish. Newmark et al. 1982: 322 says that *de* adds "intensity" to the imperative and "serves to express the speaker's impatience"; this is exactly what Greek shows, e.g., (4.9):

(4.9) a. έλα ντε 'C'mon already!'
b. σταμάτα ντε 'Stop it, OK!'

Cf. also Romani, e.g., *ava de, ava* 'I'm coming, already!'.

In Greek, ντε always occurs phrase-finally, with the exception of the fixed phrases ντε και καλά, literally '*dé* and well,' and ντε και σώνει, literally '*de* and is.enough,' that can be translated as 'once and for all!', conveying a sense of finality and annoyance.[193] In general, the emphatic *de* will be enclitic. In BSl, as noted above, *de* can be freestanding as well as enclitic, e.g., *de more de!* 'C'mon man, c'mon!,' *da de da* 'yes, of course,' and so too for the Aromanian *de*, e.g., *de bre de* 'C'mon man, c'mon,' which can also occur independently as a one-word interjectional utterance: *De!*. Romanian has the Turkish conjunction *de* in a variety of meanings, including 'and,' 'if,' and an interjection *de* meaning (roughly) 'now then; well.'

193 The use of ντε here may well be delocutive in nature, being pulled out of phrases where it signals impatience and occurs with its usual phrase-final syntax.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

In terms of meaning, the Turkish nonharmonic *de* matches the Greek, Albanian and BSl better than connector *de*/*da*, but in terms of prosodic (word order) properties, the connector *de* is the better model. It is possible of course that the postpositive connector was borrowed and simply altered in meaning in the borrowing language.[194] Boretzky & Igla 1994: s.v. also cite the use of *-ta* ~ *-da* as a focus particle or an emphatic particle after imperatives as being from the Turkish (cf. also Igla 1996: s.v.; Boretzky 1993: 87). It is clear that both Slavic and Turkish had the resources for contributing to the different usages found in the various Balkan languages, and that the two *de*'s in the two languages had the potential for various types of conflation, especially since the connective 'and' itself can serve as a kind of emphatic marker. For Romanian, Cioranescu 1958–1966: s.v. notes Moldovan *deh, dec* and mentions as well the claim that the particle is from Dacian, despite the absence of any such attestation.

### 4.3.4.2 Modifying and Expressive Discourse Elements

This last element, *de*, as noted, actually does more than just connect; it adds attitude and speaker stance, and injects a certain expressiveness or tone into the utterance, thus modifying it in some way.[195] There are also elements that are less expressive but fully modificational nonetheless. Both of these types of modifying discourse elements abound in conversation, and they have spread quite widely around the Balkans.

#### 4.3.4.2.1 Modifiers

There is a rather large class of modificational words, mostly but not exclusively from Turkish and mostly, but not exclusively, adverbs, that are borrowed into various of the Balkan languages that have something to do, in a rough way, with the evaluation of the truthfulness of the content of an utterance, offering meanings such as 'really; is it so' (thus confirming), 'certainly,' 'probably,' 'presumably,' 'perhaps' (thus commenting on likelihood), 'so to say,' 'supposedly,' 'that is,' 'as if,' 'at least' (thus mitigating or clarifying); we give several here (listed alphabetically by spelling in source language), along with their source, meaning (in donor and/or borrower, as relevant),[196] and language distribution (see Table 4.12).[197]

194 An alternative account of the shift of function of *de* has been suggested by Eva Csató (p.c., 2006). She says that it is not *de* in Turkish that has the "impatience" value but rather *be*, which might be from Greek μπρε (see §4.3.5). In other Balkan languages, however, e.g., Macedonian, *bre* and *be* are not synonymous and have distinct discourse functions. Be that as it may, she thinks, the stylistic/register-value of the one language in the other language's "home turf" may be lower, so that the Greek item in the Turkish domain is the one to be used for annoyance whereas the Turkish in the Greek domain is the one that is thus used.

195 See also §§4.3.4.2.2, 4.3.4.3, 4.3.6, 4.3.7, and 4.3.8, as well as §5.7, on expressive vocabulary more generally, not just discourse-linked elements.

196 If there is no shift of meaning from the source, then no meaning is given for the borrowing language form; the motivation for the meaning shifts, up and down (or even off) a scale of certainty, is no doubt interesting but is beyond our scope here.

197 Sources drawn on in compiling this list include Atanasov 2002, Boretzky & Igla 1994, Varol Bornes 2008, Bunis 1999, Bufli & Rocchi 2021,Capidan 1935, Cuvata 2006, 2009, Grannes et al.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.12 *Selected borrowed Balkan modifiers*

| | | |
|---|---|---|
| Trk ácaba | 'I wonder if; oh indeed!' | Alb *axhaba*, Aro *ageaba*, Blg *adžeba*, Jud *adjaba*, Mac *adžaba*, Megl *adžaba*, OEGrk *adžiba* 'I wonder; is it so?' |
| Grk αλήθεια | 'truly? really?' | Aro *alithios* 'really, truly' |
| Trk ártık | 'now; well then; not' | Aro *artic,* 'finally,' Blg *ártăk* 'finally; really; in fact,' Megl *artîk* 'finally,' OEGrk *artık* 'anymore, only' |
| Trk bári(m) [=bārī] | 'at least; for once' | Alb *bar*/*bare(m)*/*bari*,[198] Aro & Megl *báre*/*bári*/*bárim*, BSl* *bar*/*bare*/*bárem*/*barém*/*bári*/*barí*/*bárim*/*barím*, Jud *bári*, Rmn *barem*, Grk μπαρίμ, Rmi *barem* |
| Trk belki(m) | 'perhaps, maybe' | Alb *belqim*, Aro *belchi*, BSl* *belki(m)*, Grk μπελκί(μ), Jud *belki,* Megl. *belchi* 'perhaps; probably; as if' |
| Trk değil mi | 'isn't it so?' | Alb, *dilmi* Aro *delme* 'since,' BSl *delmi*/*dilmi*/*dilma* 'isn't it?', Megl *delmi* 'since; because; after,' Rmi *dilmi* 'isn't it so' |
| Trk [h]élbet(te) | 'certainly, surely' | Alb *(h)elbet(e)*, Aro *elbet(e)* 'possibly; assuredly,' BSl* *(h)elbete*/*elbetta*/*helbette*/*helbet(t)ja*, Rmn *(h)elbet*, Megl *elbet*, OEGrk *elbet(te)* |
| Trk gālibā | 'probably, presumably' | Alb *galiba* 'perhaps,' BSl* *galiba*, OEGrk *galiba* |
| Trk gerçek | 'real; really, in truth' | Blg *gerček* |
| Trk gûya, göya | 'as if; supposedly' | Alb *gjoja*/*gjyja*, Aro *ghotaha, ghoma, ghoa, gho,* etc., Blg *gjóa*, *gjoj*[*kim*], Mac *gjoa[miti]* (BCMS *đôjā*), Megl *ghiuá*, OEGrk *γ'a* |
| Grk λοιπόν | 'so; OK, well' | Aro *lipon* |

2002, Hauge 2002, Jašar-Nasteva 2001, Nehama 1977, Polenakovikj 2007, Papahagi 1974, Ronzevalle 1911, 1912, and VAF field notes. An asterisk next to BSl means the word is also attested for BCMS in Skaljić 1966 (in some cases, mutatis mutandis, e.g., *belki[m] ~ belći[m]*), although if the orthography is significantly different, the BCMS form is given in the text. Some of the Edirne Greek forms here and further below are given in transliteration to focus on their pronunciation. Those given in Greek letters are all considered 'dialectal' in Greek. Some of these words are now evaluated as 'dialectal' in other languages, and some are now archaic or not widely known. In the case of Alb *axhaba*, which occurs in the eighteenth-century *Divan* of Nezim Frankulla (Hamiti 2008: 166), Bufli & Rocchi 2021: 10 made the decision to exclude Turkisms that are only attested in eighteenth-century *Divans* such as Frankulla's, arguing that these are code switches for artistic purposes rather than "real loans." They also excluded Turkisms from Newmark 1998 that they did not find in other sources, a practice we have not followed. On occasion we indicate stress or length when they differ from what might be expected as unmarked (but note that dialectal pronunciation can also differ). A number of these words are of Arabic or Persian origin in Turkish, but for the purposes of Balkan linguistics, they are Turkisms. This list is a sampling and not exhaustive.

198 Cited in Boretzky 1976 and Bufli & Rocchi 2021, but not in Newmark 1998.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.12 (*cont.*)

| | | |
|---|---|---|
| Alb mbase | 'perhaps; maybe' | Grk μπας και 'perhaps'[199] |
| BSl pa | 'well, so, and so' | Alb (in North Macedonia), Rmi *pa* |
| Trk sāhi(h) | 'really, truly' | Alb *sahi*, Aro *saí* 'exact,' Blg *saí* usually followed by the Turkish interrogative particle *mi* to render 'Really?', also Mac & Rmi *sajmi?*, BCMS *sahi(h)* |
| Trk samsahi | 'really really' (intensive reduplication of *sahi*) | Blg *samsai* 'obviously; indeed' |
| Trk sanki(m) | 'as if' | Aro *sanchi*, BSl *sanki[m]* 'actually; that is to say; as if,' Jud *sankyi*, Rmn *sanche/i*, OEGrk *sangim* |
| Trk sözde | 'so-called; supposed(ly)' | Blg *sjuzde* 'supposedly (indicating disbelief)'; 'as if' OEGrk *seüzde* |
| Grk τάχ' | 'as if' | Aro *taha* |
| Trk yāni | 'that is to say' | Alb *jani* 'however; namely,' OEGrk *γ'a'ni*, |
| Trk zāten (coll zāti) | 'essentially; already' | Alb *zaten* 'just exactly,' Aro *zaté*, Blg *zată(n)*, Mac *zate* 'indeed, really, exactly'; Jud *zatén* 'indeed,' OEGrk *zatın* 'naturally; also' |

In the realm of truth-evaluative borrowings, the Turkish perfect marker *-mIş*, whose auxiliary 'be' form is *imiş,* is a special case. The auxiliary *imiş* in WRT is often not reduced to clitic *-mIş* in derived tenses, which is an archaism relating to its auxiliary origin. The affix itself can be treated as a separate lexical item as in the following example from Lewis 1967: 102 *Ben mişlere muşlara pek kulak vermem* 'I *miş*.PL.DAT *muş*.PL.DAT much ear give.NEG.1SG' = 'I don't pay much attention to gossip.' In Gostivar, Turkish speakers also use *miş* as a lexical item to comment on someone else's narrative in the confirmative when the interlocutor is confirming a belief rather than something s/he knows irrefutably (VAF field notes, cf. §6.2.5). Adamou 2012a notes a similar use of *-muš* in the Romani of Greek Thrace (in Xánthi), Kyuchukov 2012 reports similar usages in Romani dialects in eastern Bulgaria, and Skopje Arli also has *imiš* in such usages (Friedman 2019b).[200]

199 Andriotis 1983: s.v. derives this from a Greek μην πας (και) 'don't you go (even)' though Meyer 1894: 69 treats the Greek as a borrowing from Albanian.

200 The Cypriot Greek sentence adverb and discourse element *miši*/*miši mu* (where *mu* is a first-person reference of the type associated with the admirative-dubitative-reported complex, see Demir 2003, Friedman 2018c, Kappler & Tsiplakou 2018) is also worth noting (and see §6.2.5.10). It marks a type of evidentiality, as in εν' πλούσιος *miši mu* 'he-is rich so-they-say' (i.e., 'this is what others say, and I don't really believe it'). As footnote 166 shows, there is an *-m* that occurs in various Balkan discourse adverbials of Turkish origin, often optionally (e.g., *belki*/*belkim* 'perhaps'), sometimes in Turkish and sometimes in the borrowing language; its prevalence is interesting but we leave for another study a full investigation of its etymology and spread – while it appears to be a possessive, and clearly was taken by Cypriot Greeks as such, it may have other origins. There is also fluctuation in some borrowed forms between *-m#* and *-n#* (e.g., Albanian and Macedonian have both *tamam* and *taman*), although Turkish has two lexical items, *tamam* 'exactly' and a provincialism *taman* 'didn't I say so,' that are fairly close in meaning and form; see also footnotes 166, 181, 203, 204, 340, as well as §4.3.7.2.1 and Table 4.13.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.13 *Borrowed Balkan sentence adverbs*

| | | |
|---|---|---|
| Trk bile | 'even; already' | Alb *bile* 'even; in fact,' Aro *bile*, Blg *biljá(m)*/*bilé(m)*, Mac *bile* (dial.), Rmi *bila(m)*/*bilim*, Jud (Istanbul) *bile* (in fixed phrase from Turkish *vallahi bile* 'strewth' (Varol Bornes 2008: 457) |
| Trk hemen[201] | 'almost, nearly' | Blg *hemen*; OEGrk εμέν |
| Trk sade | 'only' | Alb *sade*,[202] Aro *sade*, Blg *sa(a)dé*, Mac *sade,* Rmi *sáde*/*sadé* Rmn *sade*, OEGrk *sadé* |
| Trk salt | 'only' | Alb *sall(a)*/*sallde*/*sallte* (dialectal), Blg *sal*/*sált(e)*, Mac *sal,* Megl *sal*/*săl*, Rmi *saltá* |
| Trk tamam | 'just right; there you have it!'[203] | Alb *tamam/n*, Aro *tamam*/*tamamá*/*tamamaná*, Blg/Mac *tamám*/*n*, BCMS *tàmām/n, tamȁm/n* Rmn (dialectal) *taman*, Grk (dialectal) ταμάμ(ι), Megl *tamam*/*n*, Rmi *tamami* |
| Rmn mai | 'almost' | Blg *mai* (Banfi 1985: 100) |

We can also observe that this reanalysis of an inflection unit as a lexical item is one of many counterexamples to the unidirectionality hypothesis of grammaticalization theory, since we clearly have a grammatical affix turning into a freestanding lexical item (see footnote 151).

Finally, numerous other elements, generally sentence adverbs, some of which have mitigating, intensifying, or focalizing discourse uses, are borrowed into other languages, mostly from Turkish though we cite one case from Romanian. As this is a more open-ended sort of borrowing, in that the definitional boundaries for such discourse elements are not fixed, we mention just a few of the more prominent ones here, and refer the reader to Grannes et al. 2002, Jašar-Nasteva 2001: 115–124, and Hauge 2002 for more examples. See Table 4.13.

#### 4.3.4.2.2 Expressives

Expressives, as noted above, introduce tone or attitude into a conversation, but we include here the borrowing of words conveying conversational pleasantries, e.g., greetings,[204] inasmuch as they inject a friendly tone. Regarding the latter, one can note the borrowing of Turkish *merhaba* 'hello' into Ottoman-era Edirne Greek in that form and Bulgarian colloquial usage as *maraba* (reflecting

201 This can be doubled in Turkish and in the various languages, with the same meaning.

202 This is given in Meyer 1891 with the adverbial meaning 'only; just'; Newmark 1998 lists it only in the sense of 'prepared without mixing; unadulterated' (as with coffee, for instance).

203 There is a form *taman* in Turkish, cited in Redhouse 1968 and labeled there as "provincial," that means 'you know; well!'. It presumably is the same word as *tamam* with something affecting final nasals, as seen elsewhere with some of these well-traveled items (see footnotes 166, 181, 201, 340). The different forms outside of Turkish may represent borrowings from one or the other of these Turkish words, though something else is perhaps going on in the individual languages; there are not enough clues from the glosses given in various sources to decide the question. See Joseph 2020d for discussion of the many variants of *tamam* found in the Balkans.

204 And see §4.3.10.1.2.2 on phraseological parallels in greetings.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

a Turkish dialect form).[205] Moreover, Aromanian in Greece has borrowed Greek γεια σου 'hello' (lit., 'health to-you' or 'health your'), given as *yeásu* by Vrabie 2000: s.v. and *yeásu* by Papahagi 1974: s.v. The Turkish formula for 'welcome,' *hoş geldin* (cf. (4.37)), occurs, e.g., as a codeswitch in Albanian epic poetry (Halimi 1951: 225) in the form *hoshgjelldën*, which reflects the WRT backing of high front vowels in closed final syllables. (Cf. also *hozhgjelden* in songs collected by Milman Parry and Albert Lord in the 1930s in northeastern Albania, Scaldaferi 2021: passim.) A similar borrowing is found in nineteenth-century Macedonian folktales as *Oždždldi, ožbulduk* (Cepenkov 1972a: 183, cited in Friedman 1995b). In this Macedonian rendering, the speaker (a Macedonian woman) uses both the greeting and the response (lit., 'welcome, well found'; cf. example 4.38) as a single greeting reflecting both local WRT pronunciation, and a transformation of the Turkish formula. Romani, too, makes use of this codeswitch representing WRT phonology and a reanalysis of the traditional answer. In a tale from Skopje narrated by a man born there in 1896, a king (*padishah*) addresses Tilči bey (a fox) "*hoš gêldın*" 'well have you come' and the fox replies "*hoš buldun*" 'well have you found,' which would be 2SG rather than the normal 1SG *buldum* 'I have found' (Cech et al. 2009: 220); on final *m* ~ *n* variation, see footnotes 166, 181, 201, 204, and 340.

A standard Albanian greeting *tungjatjeta* (lit., 'may your life be prolonged') also occurs in Macedonian folktales (Cepenkov 1972a: 120, cited in Friedman 1995b). Given the contexts of the epics and tales, it can be argued that these greetings represent codeswitching insertions determined by the addressee rather than borrowings, but the point here is that they are part of the linguistic repertoire of the narrators and their listeners. For bidding farewell, Edirne Greek used *urular olsun*, from the Turkish *uğurlar olsun* 'good luck! good journey' (lit., 'good.omens may.there.be').

Relevant in this regard, too, are the hypocoristic terms of familiar address, terms of endearment (see also §4.3.8). For instance, there are several that passed from Turkish into various Balkan languages, e.g., OEGrk ογλούμ (Ronzevalle 1911: 103), Rmi *olum* / *oglum* 'my son' (Cepenkov cited in Friedman 1995b), as well as Bulgarian *olum, jolum* (Grannes 2002: s.v.), from Turkish *oğlum* 'my boy, my son,' used as an endearment or for consolation (cf. English *my dear boy*). Macedonian also has *olum* as an archaism (Jašar-Nasteva 2001: s.v.). Similarly, Alb *xhanëm*, Mac *džanam*, Blg *džanăm*, BCMS *džanum*, Aro *gianãm, gianîm* (also *ğeanăm, ğeanîm*, Papahagi 1974: s.v.), Jud and OEGrk *džanım* (Varol Bornes 2008: 353; Ronzevalle 1911: 284) 'my dear, my soul, my dear fellow' < Trk *can-ım* (lit., 'soul-my'), is used with the same nuances in all the Balkan languages, which range from endearment to exasperation, depending on context.[206] Papahagi 1908: 163 reports expressions involving birds used for 'my dear': Alb *zogu im*, Grk πουλί μου (both 'bird my'), and BSl *pilence*, Aro *puĭlŭ*, Rmn *puiule* (all

205 Cf. Friedman 2003b on the use of Turkish *merhaba* in Gora (southwestern Kosovo) as a neutral greeting, when one is unsure whether the addressee is Goran or Albanian (cf. also Daniel et al. 2021 on Turkic and other linguae francae in Daghestan).

206 Albanian also uses *xhan* in the expression *të kam xhan* 'I love you' ('you.ACC have.1SG soul').

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

diminutives, 'little chick'). Note also Aromanian *bir* 'brave child!,' from Albanian *bir* 'son,' described by Papahagi 1974: s.v. as being "used as a term of endearment."[207] See also §4.3.8 for other such uses.

A marker that sometimes has a more challenging tone is Macedonian *demek*, from Turkish *demek* 'that is to say'; besides the connective use discussed above (see §4.3.4.1), it can also have the sense of 'really, oh yeah' (often standing alone after the utterance it is commenting on, cf. English 'as if!'). Further, Newmark 1998: s.v. describes a similar value for Albanian *demek*, saying that it "expresses disparaging doubt with irony or surprise: oh, really?," and further notes as well a use as a "parenthetical expression referring to something previous: okay, then, so." These elements could in principle be considered along with the evaluative modifiers discussed in §4.3.4.2.1, but they are included here as they seem to convey greater emotion. Adding a tone of surprise also is Bulgarian *zer*, from a Turkish source, described as follows by Grannes et al. 2002: s.v.: "question particle indicating a degree of surprise; is that so?" This most likely comes from Turkish *zahir* 'apparently, clearly, evidently,' used interrogatively.[208] Macedonian *zar* has the same use and presumably the same source. Bulgarian also has *če* indicating wonder and surprise, which may be from Romanian *ce* 'what?' in its exclamative use, i.e., 'What?!' (Hauge 2002).[209] Further, the polyfunctional expressive *ha* in Turkish, which expresses agreement, surprise, emphasis, threat, or interrogation, asks for confirmation, and (when connecting two imperatives of the same verb) marks a speaker's view of actions as going "on and on, in a burdensome way," depending on context (Redhouse 1968: s.v.), may be the source of a number of Balkan discourse expressives, especially Aromanian *ha*, described as "interjection which expresses different sentiments" (Papahagi 1974: s.v.), Bulgarian *ha* 'idem,' and possibly Greek α marking 'astonishment' (Householder et al. 1964: 139).[210]

A Turkish expressive that seems to have shifted somewhat in value in some of the languages it has entered is *gidi*. As far as contemporary Turkish is concerned, *gidi* occurs in exclamations with *hey*, referring nostalgically to the past, as in *hey gidi gençlik*! 'Oh for the days of youth!' and *hey gidi hey* 'O those times!,' and with accusative *seni* 'you,' as a term of abuse in *seni gidi* 'you little rascal.' The same usage occurs in Macedonian, e.g., in the song *ey gidi ludi mladi godini* 'O (my) madcap young years.' The function is more discourse-expressive 'expressing disapproval, threat (seriously or in jest)' and with various interjections, e.g., *ai* or *ax* 'expressing pity' (Grannes et al. 2002: s.v.; Jašar-Nasteva 2001: 123–124). Albanian *gjidi* is also from *gidi*, but it is dialectal (Boretzky 1976: s.v.) and means 'away!,'

207 And see §4.3.1.8 for more on *bir*.

208 In Grannes et al. 2002: s.v. one finds *zer* is from dialectal Turkish *zere* (standard Turkish *zira* 'because, since'), but *zahir* seems to be a better fit semantically. See above, §4.3.4.2.1, for an affirmative use of *zer* (< *zahir*), and §4.3.3.4 for a more certain outcome of *zere*/*zira* in Bulgarian.

209 Conceivably, Albanian *qe* meaning 'look! there you are! there!' belongs with *če/ce* though without an obvious sense of surprise.

210 See §4.2.4.1 regarding *ha* as a disjunctive connective, and §4.3.4.3.2 regarding *ha* as an exhortative. Universality as a source of Greek α cannot be ruled out.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

apparently in an exclamatory sense (i.e., 'Get away from here/me!'). Macedonian *gitla* is also used for 'scram!' (cf. Turkish *git* 'go away!').

Thus, these expressive items have varied origins, sometimes arising from words with lexical content that have been transferred to uses that are more discourse-based. Nonetheless, they have all become conventionalized into functions that are clearly conversational in nature. And, their conversational basis provides the conduit for their diffusion from one language into others.

### 4.3.4.3 Interjections

Since the classification of discourse markers adopted here is necessarily somewhat arbitrary, several interjections have already been mentioned, such as the affirmation markers of §4.3.4.1,[211] and the exclamatory utterances with expressive value of §4.3.4.2.2. Still, there are others, and we present here first some miscellaneous cases involving exclamations, and then focus on two types that are well instantiated in the Balkans and allow for greater depth of analysis: attention-getting words and exhortative words.

#### 4.3.4.3.1 Exclamations

Exclamatory interjections, signaling an emotive reaction to an event or development one becomes aware of, spread around the Balkans. Some are regular words expropriated for exclamatory use, while others are more on the order of noises that come to be conventionalized. For instance, for 'oops!' or 'oh!' or 'up!' or the like, one finds *hopa* in Albanian, ώπα in Greek, *opa* in Macedonian, *hop* in Bulgarian, and *(h)op* in Aromanian, And these probably should not be separated from Turkish *hop* 'now then! up! jump!' (cf. Trk *hoplamak* 'jump about, get excited,' also *hoppala* 'upsy-daisy,' etc.). Similarly, for 'alas,' Albanian and Greek both have *pa, pa, pa/*πα πα πα, and Greek has πο-πο-(πο) to signal amazement, reminiscent of Albanian interjectional *po*[212] 'oh say! But say!'; Albanian also has *bo bo (bo)* for amazement or dismay. The Albanian interjection is used by some Macedonian speakers as well, especially those with regular contact with Albanian. While these are conceivably just independently arrived at pairings of form and meaning, the clusterings are suggestive of a contact explanation, perhaps showing mutual reinforcement in conversational use.[213] BSl *lele* 'oh dear!, oh woe!' also occurs in Aromanian. One clear case with a wide distribution is *aman*, the ordinary Turkish word for 'mercy' (borrowed from Arabic) but which is used interjectionally in Turkish for 'mercy! Oh my goodness! Oh my!,' and that usage is found in

211 And note also the borrowed negation interjections in §4.3.3.3 and the other convergent uses discussed in §7.6.

212 This may or may not be connected with the *po* signalling affirmation (the 'yes' word); see Joseph 2011b for discussion of the etymology of affirmative *po*.

213 It is interesting that Ancient Greek had an interjection ὤ πόποι for 'wow; oh my,' though without information on Albanian interjections in prehistory, this fact alone does not point to Greek as the source here.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Albanian, Balkan Romance, Balkan Slavic, Greek, Judezmo, and Romani. The form *ba*, said to have originated in Greek, where it is however a discourse negator (see §4.3.3.3), is found in Turkish (Redhouse 1968: s.v.), Aromanian (Papahagi 1974: s.v.), and Judezmo (of Thessaloniki, Symeonidis 2002: 207) in the expression of surprise. In Balkan Romance, Balkan Slavic, and Romani, *ba* at the beginning of a sentence expresses disagreement or cautious agreement.

Borrowed elements figure in the expression of wishes, or a negative counterpart, the expression of wistful regret (see also §4.3.3.3 on prohibitives, and §7.6.2 on the borrowed Turkish prohibitive *sakın*) as well as approbation or disapproval. Complementizer-like, and thus functionally shifted, uses of the Greek wish-introducer 'would that . . . !' are discussed above in §4.3.3.4, but purely exclamatory borrowing is seen in Turkish *keşke* (learnèd *kâşki* from Persian *kāš ki* 'Would to God that'), 'would that . . . /if only . . . ' that is the source of colloquial BSl and Aro *kešk*i/*kešk*e 'idem' used in expressing regret (Derebej & Filipov 2019: s.v.; Grannes et al. 2002: s.v.; Polenakovikj 2007: s.v.; see also §6.2.4.2.8). Similarly, the Islamic expressions realized in Turkish as *inşallah* 'if God wills it/may it come to pass' (< Arbc 'may it be God's wish'), *maşallah* 'congratulations, bravo' (< Arbc 'what God wishes'), *eyvallah* 'thank Heavens' (Trk *eyi* 'good' + Arbc 'and God'), all occur in BSl (*išala, mašala, evala*), with the first two occurring in Albanian (*mashalla*, *ishalla*) and BRo; *mašala* also occurs in Judezmo. Balkan Slavic and Aromanian (Polenakovikj 2007: s.v.) borrow the Turkish interjection *aşkolsun* 'bravo!' (*aškolsun, ašcolsun*), which is generally used felicitously, but can also be used ironically.

Finally, by way of showing how borrowed interjectional items can be altered and even drastically reanalyzed, suggesting, as expected, that there is not always full bilingualism on the part of the speakers involved, consider Mac *spolajti* 'thank(s be to) you.' This is from Grk (ει)ς πολλά έτη 'to many years' (a congratulatory phrase said, for instance, at birthdays), but reanalyzed as to meaning and as to form. The -τη of the Greek neuter plural noun έτη 'years' (singular έτος) has been taken as the Macedonian second person DAT.SG pronoun, added onto an imperative (which frequently ends in *-aj* for singular verbs in Macedonian – Greek -α ε- ([a e]) in fast speech could yield [aj]). That reanalysis has spawned the use of second DAT.PL *-vi*, thus *spolajvi*, and the form can even be heard with a third-person pronoun, *spolaj-mu na Gospod* 'thanks be to God.' These developments are all well-motivated in Macedonian terms, especially when dealing with material that, due to its being a borrowing, is opaque.

#### 4.3.4.3.2 Attention-Getting Particles

There is a large set of varied attention-getting exclamations that spread widely in the Balkans. These are clearly conversational, in that one function they serve is to draw interlocutors together as they set the stage for starting a verbal exchange. A few somewhat localized attention-getters of a miscellaneous nature are discussed first, followed by some that are more widespread in their distribution.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

One involving negation, where the function seems to have been transferred across languages to affect native material is the use of the prohibitive negator as a one-word interjectional element with the negative imperative meaning 'Don't!' This is found in usages in Grk μη, Alb *mos*, Romani *ma*, Mac *nemoj* (Lower Vardar *nim*, Blg *nedej*, etc.; cf. R. Greenberg 1996b). Joseph 2002b speculates that this may be a calque from a Balkan Romance source, as that is a language where the same word, *nu*, is used for (independent discourse negator) 'no,' for (grammatical negator) 'not,' and for prohibitive negation, unlike these other languages (e.g., Greek has όχι, δεν/μην, and μη, Albanian has *jo, nuk ~ s', mos* in those uses, respectively, but BSl uses *ne* and Romani uses *na* as both 'no' and 'not').[214]

On the positive side of getting someone's attention, there is the Turkish presentational *işte* 'look!, here!,' which occurs in Bulgarian (Grannes et al. 2002: s.v.), Macedonian (Jašar-Nasteva 2001: 233), and Greek of Ottoman-era Edirne (Ronzevalle 1911: 98) as such, and in an apocopated form *shte* in Albanian around Elbasan and in Dibra (Çabej 2006: 80). A form that is likely native to Albanian, the interjection *xa* ([dza], cf. Mann 1948: s.v.) meaning 'here you are!,' has spread into other languages. The attention-getting here is via offering something for the taking, and this is a clue to its etymology. *xa* can be taken as an old imperative *xë* with an incorporated weak object pronoun (thus, **xë e => xa*), from the PIE root **g$^{w}$hen-*, which means 'strike, kill' in most Indo-European languages but 'hunt' in Slavic and *ex hypothesi* originally 'take' for Albanian (Eric Hamp, p.c.).[215] This *xa* is the likely source for Greek τζα, used to signal one's unexpected appearance, e.g., at someone's door, and also dialectally for revealing (presenting) oneself in the game peek-a-boo (thus "here I am!") after covering the face, a usage also found in Macedonian, where the form is *[d]za!* or *[d]ze!*.[216] It also occurs in Aromanian, as *dza*, glossed (by Papahagi 1974: s.v.) as 'an interjection by which one expresses someone's silence,' as in 'he did not utter even a *dza*!,' a use likely derivative from the presentational sense seen in Albanian and Greek (i.e., 'he did not even utter a "Here I am!"'). Cuvata 2006: s.v. gives exactly the peek-a-boo meaning for Aromanian *Dza!*; moreover, his entry (glossed by Macedonian *Dze!*) gives the explanation 'an exclamation used with small children to make them laugh: *Dza! iu-i njiclu?* 'Dze! Where is the little one?' (lit., 'where-is little.M.DEF'), so that it corresponds to English *Boo!* when used playfully between adults and children.

There are in addition two presentational words with broad instantiation in the Balkans that are likely outcomes of borrowing. These are the forms *na* 'here!; take this!,' and *ya* 'now!; now then!; well!'.

The first of these, *na*, is found in Albanian, Balkan Romance, Balkan Slavic, Greek (νά), Judezmo, and Turkish. The etymology is much disputed, with many

214 See §4.3.3.3 and §7.5 for more on prohibitives and related uses in these languages.

215 Idioms such as English *take down* for killing in a hunt may offer some help regarding a semantic connection between 'hunt' and 'take.' This root is seen also in *nxënës* 'pupil' (segmented *n-xë-nës* or *n-xën-ës*), as the 'chaser-after/taker-in of (knowledge)' (Eric Hamp, p.c.).

216 See also §5.7, where *dza* is discussed in the context of expressive phonology. And see §4.3.10.1.2.2 below for more on peek-a-boo.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Greek scholars seeing a Greek origin for it,[217] and others suggesting it is of Slavic origin, given its wide distribution in East Slavic (e.g., Russian and Ukrainian) and West Slavic (e.g., Czech and Polish), deriving from a demonstrative element.[218] It could even also be an Albanian development, if from a zero-grade imperative of the PIE root **nem-* 'take; give.'[219] Admittedly, all of these proposals could be right so that *na* would have multiple origins in the Balkans, but at least some occurrences in some of the languages, e.g., Turkish[220] and Balkan Romance, would involve borrowing. Interestingly, νά is not found in Romeyka Greek, the Pontic Greek variety still spoken in eastern Turkey in the hills south of Trabzon and especially in the region of İçgöl (Ioanna Sitaridou, p.c., March 2011), a distributional fact which is consistent with taking the presence of νά to be Balkanologically significant and with taking it as a Balkan-based borrowing into Greek.[221]

The second wide-ranging attention-getter in the Balkans is *ja/ia/ya*/για, found in Albanian, Romani, Balkan Slavic, /Balkan Romance, /Turkish, and /Greek. It is especially common phrase-initially with imperatives, adding emphasis and insistence, and functioning somewhat like English 'hey,' e.g., Greek για κοίτα 'Hey, look!,' Romani *ja phen mange!* 'So say (it) to-me!,' Bulgarian *ja mi kaži* 'Hey tell me!' (lit., '*ja* me tell'). It can also occur independently in some of the languages, e.g., Albanian *Ja, rashë e vdiqa* 'Suppose I dropped dead?' (lit., 'There! I fell and died'), or with an object, e.g., Albanian *Ja një grumbull* 'There's a bunch!,' Greek για μια στιγμή 'Hey, (wait) a moment,' Aro *ia-li vini* 'Here, he's come' (Cuvata 2006: s.v.). As with *na*/νά, *ja/ia/ya*/για presents an etymological tangle. The Greek is said (Andriotis 1983: s.v.) to be from Ancient Grk εἶα, an interjection meaning 'up! away! c'mon then!,' but Romanian *ia* for some scholars (Nandriş 1961) is from the second-person singular imperative of *a lua* 'to take,' while others (Cioranescu 1958–1966: s.v.) see it as a "spontaneous creation" ("creación espontánea"), though on a par with Alb *ja* but also with Romance forms such as Sardinian *ea*. A seemingly extended form, *iacă*, may derive from Latin *ecce* 'behold' (Cioranescu *ibid.*) and there is also *iată*, 'behold! here is!,' which may be from a reduction of *ia* with *uite* 'here (is); (look) here' (Cioranescu *ibid.*). However, Arbc *ya* is unquestionably the source of Turkish *ya* (Redhouse 1968: s.v.), so the expression could be just another Turkism, or we might be dealing with multiple sources.

This element is like an interjection in drawing in a listener, but it is also somewhat grammatical in nature in that it so typically co-occurs with imperatives.

217 Hatzidakis 1889: 332 proposes that it derives from an earlier Greek ἠνί, abstracted out of ἠνίδε (= ἴδε 'behold!' + ἴδε (impv of 'see')).

218 See Joseph 1981, 1985c for a defense of the Slavic origin proposal and a consideration of the ideology behind various etymological hypotheses.

219 This root means 'give' in Greek (νέμω 'distribute') but 'take' in Germanic (Gothic *niman* 'take' – see Joseph 2000d on the semantic difference); the derivation would be from (zero-grade) **nm̥* with the development of the syllabic nasal as in *shtatë* '7' from **septm̥-ti-*. For the possibility of a zero-grade of a root in an imperative, cf. Vedic Sanskrit *ja-hi* 'strike!' (**g$^{w}$hn̥-*), Aves *z-dī* 'be!' (**H$_{1}$s-*).

220 For Turkish, however, Tietze 2018: s.v. labels *na* 'slang' (*argo*) and connects it with earlier Turkish *nah* 'idem.'

221 See §7.8.1.2 on the syntax of νά in relation to weak object pronouns.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

In that way too, it serves a connective function in discourse, in essence announcing what is coming as a command, literally something that commands the hearer's attention.

#### 4.3.4.3.3 Exhortatives

Both *na* and *ja* are actually rather close to exhortatives, as indicated by translations given in some of the dictionaries, e.g., 'now then, come' (Levitchi 1973), 'C'mon!' (TRMJ 2005), 'here! Take this!' (Newmark 1998), among others. There are other elements like these, e.g., Turkish *ha* (see also §4.3.4.2.2), found in Bulgarian functioning as "a call to action" (Grannes et al. 2002: s.v.) and in Ottoman-era Edirne Greek (Ronzevalle 1911, 1912); note also Alb *hë*, which 'encourages action' (Newmark 1998: s.v), Rmi *ha dža!* 'get going' (lit., '*ha* + go.IMPV,' Boretzky & Igla 1994: s.v.), and Macedonian *a* (with the normal loss of etymological /h/) glossed as 'let's' (Murgoski 2013: s.v.).[222] Further, there is *éla*, formally the imperative of 'come' in Greek (έλα, from the Ancient Greek verb ἐλαύνω 'drive; sail' (Andriotis 1983:s.v.)), but used to urge people on, like English *c'mon* and borrowed into Aromanian, Bulgarian, and Macedonian in that use (R. Greenberg 1996b).[223]

The exhortative with the widest spread in the Balkans is a word meaning 'C'mon; let's go; all right already,' that in what can be called its "generic" – as perhaps the most frequently encountered – form is *[h]ajde*.[224] This seems to be best taken as deriving from the Turkish exhortative *haydi* 'hurry up! go on! all right!,' which has a number of variant forms in Turkish itself, such as *hayde*, *hayda*, and *hadi*, and derives (Tietze 2016a: s.v.) from an interjection *ha*, which has a variant *hay*, with a deictic element. Not surprisingly, there is actually quite a variety of shapes for this exhortative across the Balkans, all sharing a common core

222 Rmn *ha* expresses 'surprise' or 'interrogation' (versus *hâ*, which is 'affirmation'; cf. colloquial Turkish nasalized *hɛ̃ⁿ*).

223 There are other forms in the Balkans which have been linked with έλα, though the connections are far from certain. Albanian also has an exhortative/imperative *eja* 'come!' which Meyer 1891: s.v. took to be a borrowing from Greek έλα, but the development of -λ- ([-*l*-]) to -*j*- is hard to explain, even for a borrowing, as there does not seem to be any particular palatal quality associated with the -λ- in the putative Greek source. The unlikelihood of such a development may also stand in the way of the tentative suggestion made by Meyer 1891 that the Albanian form ultimately belongs with Turkish *ala*, a provincial variant of *elâ*, listed as a poetic "vocative exclamation" borrowed from Arabic (Redhouse 1968: s.v.) and also an exclamation (presumably a different word – not marked as from Arabic in Redhouse) meaning 'Get on!' Albanian *eja* may instead reflect the PIE root **H₁ei-* 'go,' perhaps with a particle following (cf. Joseph 2015a). Romani (e.g., in Agía Varvára in Greece) has *ale*, meaning 'hey, come on' (Messing 1988: s.v.), that appears to belong with these forms despite the seeming metathesis of the vocalism. But given that Romani 'come!' is *av!*, there are other possibilities for the explanation.

224 *Hajde* occurs outside of what we take to be the limits of sprachbund, e.g., in Croatian; this extended distribution does not vitiate in any way the value of this word, or others which occur in the sprachbund and outside as well, since the localized view that we take of the spread of individual features means of course that any given element (a word, a construction, or the like) can spread to other languages. No single element defines the sprachbund.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.14 *A widespread Balkan exhortative*

| | |
|---|---|
| hajde | Albanian (pl. *haideni*), Aromanian, Bulgarian, Romanian (spelled *haide*), |
| [de] | Greek (dialectal), Judezmo, Romani, BCMS, Slovene |
| hajdi de | Aromanian, Greek (Ottoman-era Edirne) |
| ajde [de] | Greek (spelled άιντε/άϊντε), Judezmo (Bunis 1999: 429), Macedonian (pl *ajdete*), Aromanian, Romanian (spelled <*aide*> (pl. *ajdeți*)) |
| hadi | Greek (Ottoman-era Edirne), Romani, Turkish |
| hade | Greek (Ottoman-era Edirne), Judezmo (Altabev 2003: 204) |
| ade | Greek (spelled άντε), Judezmo (Altabev 2003: 204) |
| aida | Romanian |
| aidi | Aromanian (Papahagi, Cuvata) |
| haj | Aromanian, Romanian (spelled < hai >) |
| ajt | Macedonian, Romani |
| aj/ay | Macedonian, Greek (spelled άι), Aromanian, Turkish |

nonetheless; thus one finds also *ajde*, *hadi*, *hade*, *ade*, and even just *haj* (*hay*), *ajt*, or *aj* (*ay*) (see Table 4.14).[225]

Language-specific etymologies have been proposed for some of these, yet recognizing language contact provides the best account ultimately. For instance, Greek άιντε is said (Andriotis 1983: s.v.; Babiniotis 1998: s.v.; Dangitsis 1978: s.v.; Floros 1980: s.v.) to derive from an Ancient Greek plural imperative ἄγετε ([ágete]) of ἄγω 'drive,' but the phonological developments needed for this etymology are ad hoc. While getting άι ([ái]) out of ἄγε ([áge]) is conceivable, that is as far as it goes. A change of *g* > [j] before a front vowel is regular, but loss of [j] is not; the [i] could in principle be a contraction of [j] and [e] or, if this was a northern form originally, a raising of unstressed [e] to [i], but if the latter, the raising would have to have left the final unstressed [e] intact; moreover, there is no way to get [d] from the [t] of ἄγετε. The situation is no better with the apparent variant άντε ([ade]), which is simply referred by Andriotis (1983: s.v.) to άιντε; in this case, there is no regular, non-ad hoc path either to the absence of a reflex of [g] or to the voiced [d]. Given such problems, a Greek-internal source for the Greek forms is difficult to maintain,[226] as presumably recognized by Charalambakis 2014: s.v. άντε, who, correctly takes Greek άιντε to be a borrowing, from Turkish *haydi*.

For Slavic, Skok 1972: s.v. notes that *hájde* and its variants are used throughout South Slavic, including Kajkavian, Čakavian, and Slovene. He is unequivocal that

225 Some of these forms have wider distributions; *aj* occurs in Russian, Ukrainian, and West Slavic, and both Russian and Ukrainian have *ajda* and Ukrainian also has *hajda* as cries for driving animals. Vasmer 1986–1987: s.v. takes them to derive from a Turkic source (Tatar), and thus related to Turkish *haydi*. We have normalized the spelling; for instance, the Aromanian forms are spelled <haĭ>. In some languages, the vocative particle *de* frequently co-occurs; see also §7.8.1.2 footnote 259 regarding the position of *de*.

226 See Joseph 1985a for a discussion of the ideological motivations behind these etymological proposals by Greek scholars.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

this form is of Turkish origin, from the exclamation *hay* (see above and footnote 226) plus *de* (§4.3.4.1.2),[227] although he also adduces the verb *haydamak* 'to drive cattle.' Indeed, άι-ντέ, with two accented syllables, can be heard in Greek, suggesting that the ντε originates in the impatient connective (cf. §4.3.4.1.2), and that univerbation to άϊντε may have happened independently, at least in Greek. Other South Slavic sources, e.g., Škaljić 1966: s.v. and Knežević 1962: 138 look to Turkish *haydi*. The absence of [h-] in most of Greek indicates that the [h-] forms of Balkan Romance, South Slavic (except dialects where /h/ is lost), dialectal Greek, and Albanian are from *haydi* while the initial part of Greek άιντε might be from Turkish *áy* (see Table 4.14 and footnote 226). Whatever the precise etymological connections, Turkish is primarily responsible here for the spread of this (these) exhortative(s), given the convergence in form and function. And, their highly colloquial nature confirms the importance of focusing on conversational interactions for the diffusion of such forms.

As an aside, but an interesting one that shows more evidence of how loanwords can move away from their source language features, in some of the languages, *(h)ajde*, though not an imperative etymologically, nonetheless is treated grammatically like a singular imperative, spawning plural imperatival forms, with regular personal endings added on. Thus Romanian has a "2PL" *haideți* and a "1PL" *haidem* 'c'mon; gw'an; let's go,' as does Serbian (*hajdete*/*hajdemo*), and Macedonian and Albanian have 2PL forms, *ajdete* and *hajdeni*, respectively, as does Greek, αΐντετε. Further, in addition to *hayde* (Varol Bornes 2008: 431), Judezmo has a form *aydes* (Bunis 1999: 431) that appears to be *ayde* with the 2SG ending *-s* added on, though the occurrence of άιντες in Greek makes one think of the 2SG -ς of Greek itself and also the adverbial -ς that shows up dialectally, e.g., the widespread τότες 'then' from earlier (and now standard) τότε, or επειδής 'since; because' (e.g., in Greek of southern Albania) from earlier (and now standard) επειδή.[228]

Also, Bunis 1999: 627ff. gives for Judezmo *ababam* from Turkish *ha babam ha*, literally '*ha* my.father *ha*,' as a form showing emphasis and encouragement – Redhouse 1968: s.v. *ha* glosses this as 'push on; on with you; get on.'[229] Moreover, this expression was well known in Bulgaria in the nineteenth century and is used by Konstantinov's 1895 literary creation *Bai Ganyo* (Friedman 2010a), it also occurs in the Turkish popular song sung by many Armenian musicians, *Martinim omuzumda* 'My rifle on my shoulder.'

Some discourse markers, therefore, are etymological imperatives or are treated like imperatives. Since imperatives are typically directed at second-person

227 Skok 1972: s.v. also mentions *hādı*, which seems not to be recorded in other Turkish lexicographical sources, e.g., Redhouse 1968, Tietze 2016a, TDK 1963–1977, TDK 1963–1982.

228 A final *–s* occurs in some varieties of Spanish on similar forms. For instance, Mexican Spanish has *oyes* as opposed to *oye* 'hey' (from an informal imperative of *oir* 'hear') and Valencian Spanish has *ves* 'go!,' versus *ve* in other dialects. This makes the Judezmo-internal explanation of *aydes* more likely, and raises the possibility that Greek άιντες is from Judezmo.

229 Papahagi 1974: s.v. also has *ha!* for Aromanian as an encouraging expression of the type found in Turkish.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

referents,[230] they are often accompanied by vocatives, which are terms of address also directed at second persons. Consequently, there are etymological vocatives that function as attention-getting interjectional discourse-linked elements, and these have spread in the Balkans. We defer discussion of these until the next section.

### 4.3.5 Lexical Vocatives and Related Elements

The class of conversationally based lexical items that have spread in the Balkans would not be complete without a consideration of what might be viewed as discourse-related elements par excellence, namely vocatives and other such elements. Vocatives, inasmuch as they typically involve calls to someone rather than addresses to inanimate objects, are inherently tied to conversation and to interpersonal interactions. As such they clearly fit as potential ERIC loans, and there is good evidence of sprachbund-related phenomena involving vocatives. These developments are covered in their morphosyntactic dimension in §6.1.1.4, but mention can be made here of a few relevant contact-related lexical developments.

In Judezmo of Istanbul (Varol Bornes 2008: 350, 388–389), lexical forms that serve as terms of address have been borrowed by some speakers from Greek, e.g., *kirio*/*kiria* 'sir/madame,' from κύριος/κυρία,[231] or *kukla*, for familiar address, though with a hint of an ironic sense, from Greek κούκλα 'doll.' Both Slavic and Romance speakers in the Balkans, especially in the nineteenth century, sometimes affected the title *kir* (F *kirja*), and in literature such affectations featured in comedies parodying hellenizers, e.g., Jovan Sterija Popović's comedy *Kir Janja* (Serbian) or the 1909 novella *Kir Ianulea* (Romanian) by Ion Luca Caragiale. Nevertheless, these comedies are indicative of certain social trends of their time that were reflected in language (cf. in this respect Detrez 2003 on *Gudila*s of nineteenth-century Plovdiv). Likewise, Turkish *efendi* and *hanım* were used as equivalents of 'sir/madam' or 'Mr./Ms.' throughout the Ottoman Empire (cf. also §4.3.8).

Some such borrowings look somewhat more grammatical in nature. In Albanian, for some nouns, there are special vocative forms that end in *-o*, such as *biro* 'O son!' (from *bir*) or *Agimo* 'O Agim!' (from the proper name *Agim*). This *-o* is generally assumed to be from the Slavic vocative, the one case form that remains throughout Balkan Slavic even in the regions with total loss of other substantival cases. This *-o* in Slavic is characteristic of *a*-stem nouns, most of which are feminine, e.g., *sestro* 'sister!', *ženo* 'woman!', but also *vladiko!* 'O bishop!'. In Albanian, however, the usage is generally limited to a few lexical items such as *biro*.

230 We recognize that there can be first- and third-person imperatives, at least in the sense of there being forms that fill paradigmatic cells along with second-person forms.

231 It is interesting that the actual Greek vocative form of 'sir,' κύριε, is not the form that is borrowed; rather *kirio* appears to be based on the Greek nominative κύριος. For most masculine nouns, other than *o*-stems like κύριος, the vocative singular has the same stem as the nominative but lacks the nominative marker -ς.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Further, there are occurrences of Slavic *-o* with native Romani kinship terms, e.g., Kalderash *dej* 'mother' has Slavic-influenced *dejo!* as well as suppletive *mamo!* (from Slavic) 'O mother!' (Boretzky 1994: 235), and Bugurdži *bibí* 'aunt' has both native *bibíje* and Slavic-influenced *bíbo!* 'auntie' (Boretzky 1993: 35). The lexical item *mamo* is found in both North and South Vlax dialects (Fennesz-Juhasz et al. 2003: 132, 238, 242, 248, 250). Similarly, the development of *kak* 'father's brother' into *kako* in Arli (and *kakos* in Bugurdži) seems to be a reinterpretation of a Slavic-influenced vocative into a new nominative.[232] Likewise, Meglenoromanian has *popi* 'O priest,' and *tati* 'O daddy' from Macedonian *pope!, tate!* respectively, with vowel reduction. (Atanasov 1990: 195).[233] Moreover, the Albanian vocative marker, preposed stressed *O*, is used by some Macedonian speakers.

Thus some vocative markers appear to have been borrowed in the Balkans. Still, what is perhaps the most striking Balkan lexical development pertaining to the vocative is the borrowing and ultimate spread all across the languages of the region of what has been termed an "unceremonious mode of address or cry of surprise, impatience, etc." or "exclamation [meaning] 'hey you!; you there!; well!; just!'," to use the definitions found in two Greek dictionaries (Pring 1965: s.v. and Stavropoulos 1988: s.v., respectively) for one of the representatives of this item, the Greek interjection/exclamation βρε. This and related forms seem to have originated in Greek and to have entered all of the Balkan languages, an account given by Sandfeld 1930: 20. This particle of address takes on numerous forms,[234] several of which, including some that are dialectally restricted, are listed below from the various languages (see Table 4.15).[235]

Two key facts make it clear that Greek is the ultimate source here. First, these forms have a clear etymology in Greek, being readily traceable in Greek to an Ancient Greek source – the vocative μωρέ of the adjective μωρός 'dull, sluggish,

Table 4.15 *A widely diffused Balkan particle of address*

| | |
|---|---|
| Turkish: | be, bire, bre, mari, more, mori, vre |
| Albanian: | bre, mor, more, mori, moj, mre, o, or, ore, ori, vore, vre |
| Bulgarian: | be, bre, ma, mari, more, mori, vre |
| Macedonian: | abre, be, bre, more, mori, mor', or', ore, ori, vre, |
| Aromanian: | are, avre, bre, măi, moĭ, móre, morì, omoĭ, óre, oré, re, vre |
| Romanian: | bre, mă, măi, măre, mări, vre |
| Judezmo: | abre, bre, vre |
| Romani: | be, bre, mo, mori, ore |

232 The use of *-o* for the vocative of foreign borrowings in unstressed *-a,* e.g., *tétka* > *tétko*, although it may have entered via particular lexical items, has been regularized (Boretzky & Igla 1994: 379); see §6.1.1.4.5.

233 See §6.1.1.4 on the Slavic vocative particle *le* in relation to Balkan Romance and to Romani.

234 Some of these are restricted to female addressees; see (4.10b) below.

235 BCMS shows *more*, *mori*, *bre, vre*, and in Edirne, in Ottoman times, the form βρε spread into Armenian (Ronzevalle 1912: 264). Outside of the Balkans proper, one can also cite Ukrainian *bre*, Polish *bre*, and Venetian *morè*, *bre*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

foolish, stupid, idiotic' – for most of the Greek forms.[236] Second, the greatest variety in form is to be found in Greek, with some fifty-eight distinct forms evident if one takes all of the Greek dialects into consideration (Joseph 1997a),[237] thus allowing for an inference of precedence for Hellenophone territory as the point of diffusion in ways parallel to the historical linguistic observation of greater diversity in the source than in outlying regions.[238] The full listing, showing phonetic forms arranged alphabetically with an indication of the dialect provenance, is given below, with forms previously mentioned, where the (a) forms were originally for male addressees and the (b) forms for female addressees; details on how each arose are to be found in Joseph 1997a:

(4.10) a.

| | |
|---|---|
| abré | Thrace |
| ambré | Epirus, Greek Macedonia |
| amóre | Greek Macedonia |
| ára | Velvendos (North) |
| aré | Evia, Thessaly, Greek Macedonia, Skiathos, Central Greece |
| áre | Greek Macedonia |
| avré | Thessaly |
| bre | (widespread, in Standard Modern Greek and elsewhere) |
| embré | Propontis |
| épre | Pontic |
| ípre | Cappadocia |
| már | Paros |
| maré | Epirus, Thasos, Thessaly, Thrace, Imbros, Kalymnos, Leros, Lemnos, Mykonos, Paros, Peloponnesos, Rhodes, Samos, Skyros, Central Greece, Tinos |
| máre | Mykonos, Pontic |
| mári | Thessaly, Greek Macedonia |
| maró | Vithynia |
| mbré | Greek Macedonia |
| mór | Zakynthos, Thessaly, Peloponnesos |
| moré | (widespread, in Standard Modern Greek and elsewhere) |
| móre | Epirus, Pontic |
| morés | Zakynthos, Corfu, Greek Macedonia |
| múr | Syros |
| muré | northern dialects, including Lesbos; Karpathos, Rhodes |

236 This has been well accepted (see Andriotis 1983: s.v. and Floros 1980: s.v.) since Hatzidakis 1895, who reaffirmed with more detailed argumentation a suggestion apparently first made by Koraís 1828–1835: V.33–34. μωρός (Attic μῶρος) had in Attic a feminine (nominative) μῶρος (though an ᾱ-stem feminine is apparently attested as well for Attic, through the accusative μῶραν in Herodas 5.17, if the reading is accurate). Note that Meyer 1891: 286 derives the #β- forms (especially βρε) from the imperative εύρέ of AGrk εὑρίσκω 'find (out)' (ModGrk βρίσκω, with imperative βρες), citing the meaning 'see!' from Otranto Greek. This cannot be ruled out, and other sources for some of the forms can be entertained, as the discussion below of (α)ρε as being possibly Indic in origin shows.

237 There are also two forms, βαρέ and βρ, that might be taken as further morpholexical variants but more likely than not simply reflect sandhi variants, as argued in Joseph 1997a: footnote 5.

238 For instance, there is greater dialect diversity in a smaller area in England than in the United States, and in New England and the east of the United States than in the later-populated west; also all but one of the branches of Austronesian are found exclusively on Taiwan, whence the one widespread branch, Malayo-Polynesian, originated.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

| | | |
|---|---|---|
| | murés | Greek Macedonia |
| | ór | Epirus, Thessaly, Naxos |
| | óra | Central Greece |
| | oré | Corfu, Cyprus, Crete, Peloponnesos, Skiathos, Central Greece |
| | óre | Epirus, Rhodes, Central Greece |
| | orés | Epirus, Peloponnesos |
| | pré | Thrace, Cappadocia, Livíssi, Pontic, Rhodes, Symi, Chios |
| | ρε | (widespread, in Standard Modern Greek and elsewhere) |
| | uré | Siatista (North) |
| | vór | Megistos |
| | voré | Corfu, Kefallonia |
| | vre | (widespread, in Standard Modern Greek and elsewhere) |
| | vrés | Zakynthos |
| b. | amór | Greek Macedonia |
| | amóri | Greek Macedonia |
| | arí | Evia, Thessaly, Thrace, Skiathos, Central Greece |
| | ári | Chios |
| | maí | Kos |
| | már | Evia, Kos, Livíssi, Megistos, Mykonos |
| | marí | Vithynia, Evia, Thessaly, Thrace, Cappadocia, Lefkas, Livíssi, Greek Macedonia, Megistos, Mykonos, Paros, Propontis, Rhodes, Samos, Central Greece, Chios |
| | mó | Thrace, Greek Macedonia, Tinos |
| | mór | Cappadocia, Greek Macedonia, Tinos, Peloponnesos |
| | morí | (widespread, in Standard Modern Greek and elsewhere) |
| | mrí | Greek Macedonia |
| | mú | Samothrace |
| | muí | Samothrace |
| | murí | Northern dialects |
| | orí | Epirus, Corfu, Crete, Cyprus, Naxos |
| | prí | Rhodes |
| | rá | Cyprus |
| | úrì | Central Greece |
| | vorí | Thrace, Imbros, Corfu, Tenedos |
| | vrá | Cyprus |
| | vrí | Cyprus, Lesbos, Skyros |
| | vurí | Lesbos |

The wide-ranging distribution of some of these forms across the Balkans does not mean that all of the languages borrowed the words from Greek; it is far more likely that they spread locally from language to language, and some particular instances may have a different source. Thus, for example, the forms ρε and άρε/αρέ may have Indic roots (cf. Sanskrit *(a)re* 'interjection of calling, of astonishment, of contempt, of disrespect (as to an inferior), of anger, etc.'),[239] brought

239 As Turner 1966 makes clear, these forms occur in Sanskrit, Middle Indic (Pāli, Prakrits), and throughout the modern Indic languages of India (Sindhi, Kumauni, Nepali, Assamese, Bengali, Oriya, Hindi, Gujarati, Marathi, among others). From a formal standpoint, *aré* is the vocative of the noun *arí-* 'stranger, outsider; enemy; pious,' and ρε is a "clipping" from that or better, perhaps, a resegmentation from common doubled use *are 're*, where loss of *a-* in the second part is a regular sandhi development.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

into the Balkans via Romani (Joseph 1997a). And, the form *be* may be a Turkish alteration of μπρε, given that Turkish generally does not allow initial consonant clusters,[240] so that its occurrence in Bulgarian, Macedonian, and Romani would reflect borrowing from Turkish. Moreover, Vastenius 2011 argues that the *b*-initial forms may have a different, more easterly origin (in Kurdish *bra* 'brother'), from the *m*-initial, clearly Greek-derived, forms. Various forms have other proposed etymologies, too numerous to adduce here. Nonetheless, it is still the case that a good many of the forms seem to emanate ultimately from Greek.

These particles of address are very common in all of the languages, and in some instances have taken on uses that go beyond simple address or attention-getting. In Kefalonian Greek, for instance, μπρέ is found as a neuter noun meaning 'wife,' το μπρέ μου 'my wife' (ILNE: s.v.), and βρε and other related forms can be used as an expression of surprise or wonder (as in other Balkan languages). Further, according to Tannen & Kakava 1992: 29–30, ρε serves as "a marker of friendly disagreement" in Greek conversation, presumably a pragmatic function of its interjectional use, whereas Costanzo 2009 discusses βρε as a marker of solidarity, drawing a typological parallel with certain Chicano Spanish terms; Vastenius 2011 sees solidarity but also social power as relevant for the pragmatics of some of these forms in various of the languages.

Given their high frequency in conversation and interpersonal interactions, these forms would spread quite easily when speakers of different Balkan languages communicated with one another in their varieties of their interlocutor's language. They would be highly salient in terms of their function, and easily inserted into conversation, and once they spread, they could become entrenched, as they would form useful "bridge" words between languages, whether speakers were drawing on their own language's resources or trying to use words from their interlocutor's language.

### 4.3.6 Onomatopoeia and Related Words

While conversation is a means for conveying information, it also conveys feelings, reflecting an emotive side to human interactions (Jakobson's *phatic*, Bühler's *expressive*; see §6.1.1.4). By turning to expressiveness here, we are signaling that the intimate contact we envision in the Balkans had this kind of expressive side to it as well. Onomatopoeia, as the set of conventionalizations of noises and sounds from the natural world, most particularly but not exclusively animal noises, is one aspect of expressive phonology (treated more fully in §5.7). As such, given its expressive nature, it is not surprising to find parallels across the languages, even in the face of issues of naturalness and universality for individual noises. Nonetheless, Emeneau 1969 has determined on the basis of structural properties on the form of

240 The only words in contemporary Turkish with initial *br-* are relatively recent loanwords or are the result of rapid speech clipping, e.g., *bırak-* > colloquial or dialectal *brak* 'leave' (which itself, however, has a difficult etymology; see Tietze 2002: s.v.); note also *bire* (Table 4.15), reflecting a common type of resolution to the problem posed by an initial cluster.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

onomatopes in Dravidian and Indic, and on shared phonological material that allows for what he calls "areal etymologies," that there was contact-driven diffusion of onomatopoetic material between these two language groups, with many Dravidian features entering Indic via contact; indeed, he takes this as one of the defining characteristics of the South Asian sprachbund.

Related to onomatopoeia are words that are themselves noises but are not directly mimicking natural sounds. English *shhh* to quiet someone or *psst* to get someone's attention would be instances of such words.

As indicated above, there is a potential methodological pitfall with any attempt to establish a connection of a historical nature such as borrowing or contact-induced influence regarding onomatopes and related noise words in two or more languages. Since mimicking a natural sound is what onomatopoeia is all about, one necessarily has to worry about how to rule out naturalness and universality as the cause for any similarity across languages in this domain; given naturalness, an onomatope could have arisen independently at any time in a given language. The case of the noise made by cats is instructive in this regard. We find the following as a conventionalization for a cat noise in the Balkans (see 4.11):

(4.11)

| | |
|---|---|
| Albanian: | mjau |
| Bulgarian: | myau |
| Romanian: | miau |
| Greek: | νιάου ([niau]) |
| Turkish: | miyav, miyauv |

The initial nasal, the high front vocoid, and the diphthongal vocalic nucleus all match across the languages, but since this word sounds so much like a cat's vocalization, there is no compelling reason to insist that these forms have anything to do with one another other than reflecting a universality of the human experience with cats. Consider in this regard Tamil *miau* for the same sound. Thus it is not of particular Balkanlogical interest.

The most compelling instances that might point to contact as playing a role in the onomatope are ones where the forms in the different languages share some unnatural oddity that would point to a common origin or common or even mutual influence.[241] Working in this way, it is possible to identify some Balkan onomatopes and noise words that show the effects of contact among speakers. Some of the convergences, moreover, are quite localized and restricted to just a few languages.

Perhaps the most compelling example of a borrowed onomatope in the Balkans involves the sound of laughter. Grannes et al. 2002: 147 in their discussion of Bulgarian *kis-kis* for that sound explicitly treat it as a borrowing from Turkish, which has *kıs kıs* as well as, dialectally, *kis kis* for the same noise (specifically the kind of covert laughter denoted by English *giggle*, cf. *tee-hee*). Here the exact

241 Methodologically, this is like the "*fait particulier*" ('singular fact') that Meillet 1925 posits as being particularly revealing of common origin when one is working out reconstruction via the Comparative Method or seeking to establish genetic relatedness of languages. See Harrison 2014 for some general discussion of this methodological step.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

match in sounds, with an initial *k* and a final *s* that are not found widely cross-linguistically, and especially not in this particular combination, allows for a good case to be made for borrowing as opposed to independent origin.[242] Another such example is the conventionalization for a knocking noise: *tak-tak*/*tac-tac*/τακ-τακ occurs in Balkan Slavic, Judezmo, /Balkan Romance, and /Greek. It is likely that these are connected to the Turkish noun *tak* (underlyingly /takk-/) 'a thump, knock,' which was borrowed as a noun into Edirne Greek of the Ottoman period. Balkan Romance also has *tac* as an interjection imitating a cracking noise. In this regard, too, Edirne Greek *takır-tukur* for the sound of footfalls or of a hammer is noteworthy, as it appears to be from Turkish *takır takır* 'noise of a horse's hooves' and/or *takır tukur* 'alternation of tapping and knocking sounds,' where the reduplication as well as the segmental matching, even down to the adoption of the unrounded back vowel [ɪ], points to a borrowing. Balkan Slavic and Greek have repeated *tak-tak[-tak]* for 'knock-knock[-knock],' also a likely contact-induced convergence.

Somewhat striking too are the words for the sound of a goat bleating, since Greek and Turkish match up well, with [mæhehe] and με.ε.ε ([mɛ.ɛ.ɛ]) respectively, as two of the only three languages (Russian being the other), out of sixteen surveyed,[243] with a tri-syllabic conventionalized noise. Still, this is dangerous territory in which to draw too many conclusions.

In some instances, it is not so much a clear borrowing as some partial convergences that give a localized clustering suggestive of mutual influence. This is the case with the noise for a dog's bark (4.12):[244]

(4.12)

| | |
|---|---|
| Albanian | ham-ham |
| Aromanian | ham-ham, gap-gap |
| Bulgarian | bau-bau; džaf-džaf |
| Romanian | ham-ham; hau-hau |
| Greek | γαυ-γαυ ([γav γav]) |
| Macedonian | av-av |
| Romani | hau-hau |
| Turkish | hev-hev; hav-hav |

Although the forms are not identical, similarities are evident in the occurrence of an initial back (velar or glottal) fricative in all but Bulgarian and Macedonian, but even the absence of an initial consonant in Macedonian can be reconciled with the prevalence of [h-] in the other languages because Macedonian (or more accurately, its western dialects, on which the standard is based) historically lost [h]/[x] by

242 The initial *k* in a laughing noise is found in Korean, and in both Ancient Greek and Sanskrit the verb for 'laugh' – Grk καχάζω, Skt *kakhati* – has initial κ/*k*, representing an underlying aspirated χ/*kh* due to the effects of Grassmann's Law eliminating aspiration in the first of two successive aspirates. A form with initial *h-* is common in the Balkans and elsewhere, and seems too natural to be a matter of any sort of contact involvement in a convergence.

243 From www.eleceng.adelaide.edu.au/Personal/dabbott/animal.html, at the time the research for this section was conducted, but the link is now dead. Still, the currently active https://en.wikipedia.org/wiki/Cross-linguistic_onomatopoeias gives similar results.

244 In part from http://en.wikipedia.org/wiki/Bark_%28utterance%29#Representation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

a regular sound change (cf. *ubav* 'beautiful' vs. its cognate in Bulgarian *xubav-*, *endek* 'ditch' vs. its Turkish loan source *hendek*), so that [av av] matches up as expected with the initial *h-* forms. Moreover, while *h*-initial dog-noise words occur in other languages outside of the Balkans, e.g., Finnish (*hau hau*), various Slavic languages (Ukr [ɦaw-ɦaw], Sln *hov hov*, Pol/Slo *hau hau*, Cz *haf haf*), they are mostly located in Eastern Europe and the Eastern Mediterranean (cf. Arbc/Heb *hau hau*), or contiguous with that (e.g., Arme *haf haf*), interrupted by Croatian and Hungarian (*vau vau*). Only Thai with *hoang hoang* has #h- outside of this extended region, though there are other individual matches with some Balkan languages, e.g., Tagalog *aw aw* and Lithuanian *au au* resemble the Macedonian form. And while a final labial of some sort is frequent cross-linguistically, a final *-m* is found only in Albanian and Balkan Romance. The Balkans may thus be part of a larger zone for this sound, and there are smaller clusters (e.g., perhaps a Habsburg zone for Croatian and Hungarian), but there is a definite concentration of similarities in the languages identified herein as Balkan languages nonetheless.

Turning to noise words in the Balkans offers some particularly clear cases of borrowing in this sector of the lexicon. For instance, for attracting a cat, in Greek one says ψι ψι ψι/*psi psi psi*, like the *ps ps ps* of Balkan Slavic and Balkan Romance. There is convergence too with the noise used to quiet someone (like English *shhh* or *shush*), for Albanian has *sus*, as does Macedonian and Turkish, with Turkish being the likely source. Greek here has σους but also σουτ, matching the initial and the vocalism of *sus*, and somewhat matching the Bulgarian *št*, certainly as to its final and even in the initial once allowance is made for the absence from the Greek phonemic inventory of a palatal fricative. The preciseness of these matchings makes it reasonable to assume that contact played a role in the convergence and such examples show that there can be cross-language influence even in the expressive domain of the lexicon.

### 4.3.7 Reduplication

Reduplication, the repetition of linguistic material generally though not exclusively at the word or sub-word level, is exploited in many languages for expressive purposes;[245] Moravcsik 1978: 316–325, for instance, documents as typical for reduplication such cross-linguistically recurring functions as intensity and emphasis, and even diminution and related notions like endearment but also derogation, all of which involve the addition of emotion and color, and the like, into discourse. These functions often have an iconic basis, a characteristic that is

245 We note, though, that languages also use reduplication for various grammatical functions, such as iteration, distributivity, and duration; these do not seem to figure in Balkan usage, though the constructional reduplication found for 'whether . . . or not' expressions discussed in §4.1 might count as grammatical in some sense. We exclude here the "object reduplication" construction, discussed in §7.5.1, as it is not repetition of form but rather involves repetition of reference, with concord regarding certain grammatical features (gender, case, number).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

aligned with expressivity.[246] As expressives, instances of reduplicative word-formation are especially well-suited for use in conversation, inasmuch as human verbal interaction frequently involves emotional coloring and more than just the exchange of information, which are very common in conversation. Since reduplicative formations show diffusion throughout the Balkans, they are thus prime candidates for consideration as ERIC loans.

There are three main types of reduplication that are relevant here, two of which are noted in Asenova 2002: 276–290 and from her account fall under her rubric of "full Balkanisms" (see also §2.4 (especially footnote 37)).[247] The first, discussed in §4.3.7.1, fits in with universal characteristics of reduplication but does have some contact-related aspects to it, possibly including, though, Balkan-external ones. The other two, discussed in §4.3.7.2, are definitely contact-related as they involve demonstrably Turkish patterns found in various forms borrowed into other languages in the Balkans. Beyond these patterns, there is a single reduplicative expression that is noteworthy.

### 4.3.7.1 Whole-Word Reduplication

Seliščev 1925: 46, 51–57 appears to be the first to draw attention to the Balkan "*repetition du substantif, d'ordinaire à l'accusatif, pour exprimer des différenciations dans une quantité, des parties isolées, des groupes, des series*" ('repetition of a substantive, usually in the accusative, to express differences in a quantity, isolated parts, groups, series'), and he illustrates it with examples from Bulgarian, Albanian, Greek, and Aromanian. This phenomenon was mentioned by Sandfeld 1930: 162, referring to Seliščev. Asenova 2002: 276–290, as noted above, is the only modern handbook treatment of the Balkan languages that mentions this construct as a possible Balkanism; Asenova 1984 is replete with Greek and Bulgarian examples. A sampling is given in (4.13):[248]

(4.13) Alb copa-copa 'all in pieces' (Newmark 1998: s.v.)
copëra-copëra 'piece by piece' (Seliščev 1925: 53, from the 1802 *Tetraglosson* of Daniil Moschopolitis; see Kristophson 1974)
pika-pika 'one drop after another, drop by drop'
pjesë-pjesë 'piece by piece'
valë-valë 'wave upon wave'
vende-vende 'here and there; in several places'

246 For instance, reduplication of a word creates a combination with more of the same material in it than the individual word has; the result is iconically a more intense version of that word as there is simply more of the word evident. Such iconicity means the form has a less arbitrary connection to its meaning than is found with more ordinary vocabulary, exactly the sort of relation that expressive language often shows.

247 Asenova's approach differs somewhat from our own insofar as she includes collocations that are outside the scope of what we mean by reduplication here, e.g., positive + negated verbs of the type we discuss in §4.1.

248 We have included here some adverbial examples as well. See also footnote 251. Kallergi 2015 is a detailed study of this construction in Greek.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

| | |
|---|---|
| Aro | agalea galea ‘nonchalantly’ (Papahagi 1974: 116; base word from Grk αγάλια ‘slowly’) |
| | bucătsi bucătsi ‘piece by piece’ (Seliščev 1925: 53) |
| | pale pale ‘line by line’ (borrowed from Alb) |
| Grk | τοίχο τοίχο ‘along the wall’ |
| | γιαλό γιαλό ‘along the sea’ |
| | δύο δύο ‘two by two’ |
| | κομμάτι κομμάτι ‘piece by piece’ (LKNonline: s.v.) |
| | φιρί φιρί ‘insistently’ (from Trk *fırıl fırıl*, LKN: s.v.; Babiniotis 1998: s.v.) |
| Blg | na vălni na vălni ‘in waves’ (Sofia region, Seliščev 1925: 52) |
| | na tumbi na tumbi ‘in groups’ (Sofia region, Seliščev 1925: 52) |
| | kupove kupove ‘in heaps’ (Sandfeld 1930: 162) |
| Mac | komati komati ‘piece by piece’ (borrowed from Grk, Seliščev 1925: 53, from the 1802 *Tetraglosson* of Daniil Moschopolitis; see Kristophson 1974) |
| | kupoi kupoi ‘in heaps’ (Prilep dialect, Seliščev p. 53) |
| | na tumbi tumbi ‘in groups’ (Prilep, Seliščev p. 52)[249] |
| | (po)leka-poleka ‘slowly, little by little’ |
| Rmn | încet-încet ‘slowly-slowly, little by little’ |
| Trk | yavaş yavaş ‘slowly, gently, take it easy’ |

Seliščev observes (p. 54) that the doubled pattern occurs in Turkish (e.g., *kapı kapı* ‘door to door,’ cf. Göksel & Kerslake 2005: 100), and indeed one can find instances of this doubling in the Balkans with Turkish-derived words, as with Alb *dallgë-dallgë* ‘in waves; wavy,’ Mac *dalgi dalgi* ‘idem’ (cf. Trk *dalga dalga* ‘idem’).

Seliščev thinks that Turkish played some role in this Balkan reduplicative pattern, in particular the bare nominal doubling found in some parts of modern Balkan Slavic, such as the Macedonian *kupoi kupoi* (StMac *kupovi kupovi*) and Bulgarian *kupove kupove* noted above in (4.13), but he is quick to point out that a certain type of nonbare nominal doubling occurs in early Slavic and the bare type occurs in early Greek, thus predating contact with Turkish in each case. The Slavic Gospels have *na spody na spody* ‘by groups,’ in Mark 6: 39–40, with the preposition *na* included in the doubling, and Greek here has συμπόσια συμπόσια ‘by companies’ (6: 39) and πρασιαὶ πρασιαί ‘by groups’ (6: 40). Other examples in the New Testament, e.g., Mark 6: 7 ‘by twos,’ with δύο δύο in the Greek compared with *dъva dъva*, literally ‘two two’ in the Slavic Gospels but also *dъva i dъva* ‘two and two’ and *dъva nъ dъva* ‘two but two’ in different early (eleventh-/twelfth-century) renditions, suggest that the simple bare-element doubling in early Slavic was a calque on the Greek and that, as Seliščev (p. 56) puts it, “*le redoublement du nom de nombre avec valeur distributive répugnait au sens linguistique des Slaves*” (‘doubling of a noun of number with distributive value was repugnant to the linguistic sensibility of the Slavs’), so that something, e.g., a conjunction, needed

249 Actually, this example comes from an author, Eničerev, born in Plovdiv but educated in Prilep. Cf. also Sonnenhauser 2015 for additional examples.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

to be added to make the doubling in the translation of the Gospel more natively Slavic (hence *dъva i dъva*/*dъva nъ dъva*). And, there are instances in Greek that predate the New Testament, most importantly the Classical Greek example μῡρία μῡρία 'by the tens-of-thousands' (lit., 'ten-thousand ten-thousand') found in Aeschylus's *Persians*, l.980 (472 BCE).[250]

But far from being a Greek pattern that spread, there are indications that the doubling could simply be independent in each language, a possibility that Seliščev both is aware of and does not entirely dismiss and that Sandfeld is inclined to credit. That is, a doubling pattern is found in many languages in expressions of distributivity and what Moravcsik 1978: 318 calls "scattered plurality," e.g., Quileute, Twi, Yoruba, and Mitla Zapotec, and so it most likely reflects simply a universal iconic way of referring to multiple instantiations of essentially the same entity. Both Seliščev (p. 46) and Sandfeld (p. 162) note its occurrence in Italian, and in broader Indo-European terms, it is a regular compound type in early Sanskrit (*āmreḍita*), e.g., *dive-dive* 'day by day; daily,' *pade pade* 'in every place,' and elsewhere; thus the pattern could in principle be a parallel inheritance in each language from Proto-Indo-European and not a contact phenomenon. Indeed, Stolz

250 Seliščev (p. 56) cites two examples that predate the New Testament, but each presents some problems. He gives a μίαν μίαν ('one one,' F.SG.ACC) as being from a lost play of Sophocles, which he identifies as Ἔρις, and he says that this expression was considered vulgar, i.e., low-style, by the anti-Atticists of the Hellenistic era. We have been unable to verify this cited form, though authorities seem to think it could be real. It is given by LSJ: 492 as coming from "S. *Fr.* 201," referring to Sophocles *Fragmenta* (ed. A. C. Pearson, Cambridge 1917), though the reference to Pearson is updated, in the "Revised Supplement" to the volume (edited by P. G. W. Glare, 1996: xix), to "H. Lloyd-Jones, N. G. Wilson O[xford]C[lassical]T[exts] 1990," and the entry for μίαν μίαν is corrected (p. 106) to "r[e]p[or]t[e]d[ly] μίαν μίαν, app[arently] 'from time to time', S. *fr.* 201." Similarly, the Septuagint example he gives, εἷς εἷς, literally 'one one,' from 1 Chronicles 24: 6, is also problematic. While it does occur in a passage where a distributive sense is reasonable, since different households are being allocated to different groups, the syntax of the Greek is unusual and bears a curious relation to the Hebrew original. The Hebrew reads *bet-av 'echad 'achuz lə'elə'azar u 'achuz 'achuz lə'ithmar* 'house.of-father one being.taken for.Eleazar and being.taken being.taken for.Ithamar,' i.e., 'one principal household being taken for Eleazar and [one] taken for Ithamar,' with the passive participial form *'achuz* 'being taken' repeated, presumably to signal the distributivity of the actions, but with only a single instance of *'echad* 'one.' The Greek has nothing corresponding to the participial and has 'one' occurring doubled twice (οἴκου πάτρᾱς εἷς εἷς τῷ Ἐλεάζαρ καὶ εἷς εἷς τῷ Ἰθαμαρ, literally 'house.SG.GEN family.SG.GEN one.M.SG.NOM one.M.SG.NOM DEF.ART.M.SG.DAT Eleazar and one.M.SG.NOM one.M.SG.NOM DEF.ART.M.SG.DAT Ithamar', and moreover this group of words does not fit in syntactically with the rest of the verse. It seems as if the Greek doubling of the numeral is somehow reflecting the doubling of the participial in the Hebrew, but the forms do not really match up, neither lexically nor semantically nor syntactically. Accordingly, we consider both of these pre-NT examples as dubious, though the Aeschylus example and the NT examples are unquestionably clear and thus offer suitable early testimony for this construction in Greek. And there may well be other early instances to reckon with – Torallas-Tovar & Martín 2020 mention κατὰ δύο δύο 'two by two' from a magical papyrus. On the other hand, the Slavonic *Apocalypse of Abraham*, which was translated from a lost Greek original, itself a translation from a lost Mishnaic Hebrew text, has the Church Slavonic *pomalu pomalu* from Septuagint Greek κατὰ μικρὸν μικρόν corresponding to the Masoretic *mə'at mə'at* 'little by little, gradually' (Deuteronomy 7: 22; Kulik 2005: 13, 77). In Biblical Hebrew, *mə'at* was a noun meaning 'fewness, small amount,' so the nominal doubling is quite old for Hebrew, and if Hellenists considered it vulgar, this would be consistent with it being perceived as of "barbarian" origin in Greek.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

2006 (see also Stolz 2003) treats whole-word reduplication, what he calls "total reduplication," as a potential language universal, so that its occurrence in any given language, under this view, need not involve contact or inheritance.

Still, for the specific doubling that occurs in modern Balkan Slavic without an added element cited above, Turkish influence can reasonably be invoked to explain the absence of a conjunction. Moreover, for individual lexemes or lexical constructs, contact is certainly indicated, as the borrowed *komati komati* (Mac) and *agalea galea* (Aro) examples above show. One widespread reduplicative phrase in the Balkans is Turkish *yavaş yavaş* 'slowly, gradually' (lit., 'slow slow'), a doubling of *yavaş* 'slow, gentle, docile,' a form which has also entered these languages (see Table 4.16).

The languages also show other formations with similar meaning composed of native, or at least non-Turkish, material; see Table 4.17.

And in some instances, influence from the reduplicated pattern can be detected; while Greek σιγά 'slowly' derives from the Ancient Greek adverb σιγῇ 'in silence'

Table 4.16 *Reduplicated 'slowly' in the Balkans*

| | |
|---|---|
| Alb | avash avash 'little by little; very slowly' (cf. *avash* 'slowly; softly,' with variant *javash*) |
| Blg | javaš javaš 'at an easy pace' (cf. *javaš* 'idem') |
| Mac | javaš javaš 'slowly, take it easy' |
| Grk | γιαβάς γιαβάς ([javas javas]) 'slowly'[251] |
| Aro | iavash iavash 'slowly, take it easy' |
| Rmn | iavaş iavaş 'gently, without hurry' |
| Jud | yavaš yavaš 'very gently, without hurry' |

Table 4.17 *More reduplicated 'slowly' in the Balkans*

| | |
|---|---|
| Alb | dalë-dalë 'slowly'<br>dalë ngadalë/dalëngadalë 'unhurriedly, slowly; little by little, gradually' (*nga* 'from; out of; by') |
| Blg | léka poléka 'little by little; bit by bit' (cf. *lek* 'light', with *po-* 'more') |
| Mac | [po]leka póleka 'little by little' |
| Grk | σιγά σιγά 'slowly; little by little' (cf. σιγά 'gently, softly; slowly') |
| Aro | agalea (a)galea (from Grk αγάλια 'slowly') |
| Rmn | încet-încet 'slowly-slowly, little by little' |

251 Among the large standard lexicographical resources, Babiniotis 1998: s.v. and Charalambakis 2014: s.v. have this phrase, marked as "popular" (and for the former, "rare"); it also occurs in lyrics of popular songs (e.g., rebetika).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(a frozen dative case),[252] the earlier form occurs only singly, not doubled; hence the repetition in σιγά σιγά is likely to be a reflection of the doubling pattern with this sememe in neighboring contact languages.[253]

Moreover, the Albanian forms *dalë dalë/dalë ngadalë/dalëngadalë* may reveal a further, different contact-related detail. Although the semantics of the connection are challenging, inasmuch as adjectival *dalë* by itself means 'protruding, sticking out; worldly,' these forms appear to involve whole-word doubling. The latter two show the occurrence of a linking element, in this case a preposition, between the doubled pieces. That is, the Albanian *dalë ngadalë/dalëngadalë* presumably are etymologically **dalë nga dalë* , where *nga* is the preposition 'from' (Geg *kah*, and note Geg *kadal* 'slowly'). This use of *nga* is striking, as it agrees with what is seen in Slavic and with what, to judge from Seliščev's discussion summarized above, is needed at an early stage to make the doubling suitable for use in Slavic; it is reasonable to suppose then that the Albanian usage here reflects a contact effect and is due to Slavic influence, the result of a calquing on a Slavic model. Similarly, while Aromanian in 1802 showed *bucãtsi bucãtsi* for 'piece by piece,' as given in (4.13) above, Papahagi 1974: 289 cites the same expression as *bucãtsi di bucãtsi*, with the preposition *di* 'from' between the doubled pieces, just as in the presumed Albanian **dalë nga dalë* (cf. also Mac *leka poleka*), again a likely contact effect, either via Slavic or via Albanian.

### 4.3.7.2 Turkish-Origin Reduplication Patterns

Besides this whole-word doubling, Turkish has two reduplicative patterns that have had an impact on the Balkan lexicon. These are detailed in the two subsections that follow.

#### 4.3.7.2.1 Turkish CVC-Intensive Reduplication

Turkish is a language that in general does not have prefixes, yet it does have one prefix-type that is reduplicative in nature, serves an expressive function, and has diffused into several languages of the Balkans. This is the CVC-intensive prefix, generally used just with adjectives, which copies the initial consonant and vowel of the base word (or just the vowel if vowel-initial) and closes with another consonant, most often *p* but also *s*, *r*, or *m*. As befits an intensive, the meaning signaled is, as described by Göksel & Kerslake 2005: 98 one of "accentuating the quality of an adjective," thus it is 'very; highly' or the like. For instance, *beyaz* 'white' forms an intensive *bembeyaz* 'very white,' *eski* 'old' forms *epeski* 'very old,' and *temiz* 'clean' forms *tertemiz* 'clean as a pin.' There are, however, some exceptional forms, e.g., *çıplak* 'naked,' *çırılçıplak* 'stark naked.'

252 On the connection between silence and slowness, Turkish may also have played a role, since *yavaş* when used with a verb of speaking, e.g., *konuş* 'talk, speak,' it means 'speak in a low voice,' i.e., in near-silence. Babiniotis 1998: s.v. has a prescriptive usage note saying that the '(speak) softly' meaning is "λανθασμένη," i.e., 'erroneous.'

253 Cf. Newton 1962 and Joseph 2003c: 681 on this; while Newton mentions σιγά σιγά, he overlooks the fact that Ancient Greek here used σιγῇ only singly, not doubled.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

There are traces of this formation in various languages mainly through the borrowing of both the base adjective and the prefixed intensified adjective. In Greek, for instance, there is τσιρ-τσιπλάκης 'stark naked,' from Turkish *çırçıplak* 'idem' (this form is also acceptable, cf. above) alongside τσιπλάκης 'naked' (Trk *çıplak*), and in Bulgarian and Macedonian one can find several such pairs (Grannes 1974/1996: 139; Grannes et al. 2002; Jašar-Nasteva 2001), e.g., *bambaška* 'peculiar, strange' (Trk *bambaşka*) alongside *baška* 'different' (Trk *başka*), *samsai* 'obviously; indeed' (Trk *samsahi* 'really really') alongside *sai* 'really' (Trk *sahi* 'really'), *tastamam* (dialectal) 'perfect' (Trk *tastamam* 'complete, perfect') alongside *tamam* 'exactly, perfect' (Trk *tamam* 'complete, finished, just right, true'). Such pairs, if one were to take a purely synchronic viewpoint, would lend themselves to an analysis by which Greek and Balkan Slavic have a very limited "intensive prefixal reduplication process"; such an analysis, however, would not mean that the Turkish process itself was borrowed here, but rather that the lexical material that gives evidence of such a process was borrowed, from which the process would then be "re-created" in the borrowing languages.

This process is generally found just with Turkish lexemes that have been borrowed. However, M. Ivić 1984 has shown that in some of the languages there is a limited degree of productivity for this process, mostly with monosyllabic stems with meanings such as naked, full, alone, entire, sound, new, and similar states, and even involving native – or at least, for some languages, non-Turkish – material.[254] One widespread such case is the Slavic root *gol-* 'naked,'[255] which is the basis for the following prefixed intensives, all meaning 'stark naked': Blg *gol-goleničăk*, Mac *gol-goleničok*, and Rmn *gol-golut* (as a borrowing from Slavic), as well as BCMS *go-golest;* also *go-golcat* or *gol-golcat*.[256] We can also note here BCMS *sam-samcat* 'all alone' from native *sam* 'alone, etc.'

#### 4.3.7.2.2 Turkish *m*-Reduplication

Turkish also has a type of whole-word reduplication that replaces the initial sound of the word with *m-* and repeats the rest of the word. This is sometimes referred to in the literature by its Turkish name, *mühleme* (e.g., by Stolz 2006, for whom it is a type of "total-reduplication-*cum*-variation"), or *reduplication with m*. The meaning of the resulting composite form is a type of collective for the item itself and/or

254 Here is a suitable place to offer a correction to a suggestion made in Friedman 1986c that Macedonian *zim-zelen* 'evergreen (tree),' cited with the gloss 'bright-green,' was a lone case of this Turkish intensive reduplication with native Slavic material. Rather, this is probably made up of Macedonian-internal parts giving a compound: *zim-* from *zima* 'winter' + *zelen* 'green,' i.e., 'green in winter.' Even if a compound with native elements, however, *zim-zelen* has the structure of the Turkish pattern synchronically. We can also note here Mac *ugol gol* 'stark naked,' where *ugol* is not reduplicative but is a separate lexical element that means 'completely.'

255 Cognates elsewhere in Indo-European, e.g., Lith *gala* 'naked' and Eng *callow*, show this to be an inherited root.

256 It is interesting that this particular sememe is involved here, as it appears to have been subject to a number of interesting alterations around Indo-European, most likely due to taboo-related issues. For instance, from a likely PIE *$neg^w$-ment-/$neg^w$-no- (cf. Hitt *nekumant-*, Skt *nagna-*), Greek shows a metathesis to γυμνός (as if *$g^w$on-m. . .), and Avestan shows dissimilation to *magna-*. One can compare also the English folk-etymological (and/or taboo-related) alterations of *start-naked* ("tail-naked") to *stark-naked* and *butt-naked* to *buck-naked*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

things related to it. Göksel & Kerslake 2005: 99 characterize the function as a way "to generalize the concept denoted by a particular word or phrase to include other similar objects, events or states or affairs." Often a translation with 'and such, and the like' conveys the sense. For instance, *kitap* 'book' yields *kitap mitap* 'books and such,' *dergi* 'magazine' yields *dergi mergi* 'magazines and such,' *içecek* 'drinks' yields *içecek miçecek* 'drinks and the like,' and so on. There can be also a dismissive or minimizing sense associated with this construction, as in *yeşil meşil* 'green-ish' (cf. *yeşil* 'green'); this is often accompanied by a somewhat pejorative sense, and these features give the *m*-doubling an expressive quality that renders it very useful for, and as a result very common in, conversation.

This pattern has an interesting history, probably originating far to the east of Asia Minor, possibly South Asia or beyond, since there are parallels in modern Indic and Dravidian and even other Asian languages, and most likely it spread westward (see Southern 2005 and Levy 1980 on this) into Turkish as well as the Caucasus.[257] Thus this is a pattern that can diffuse, and it is found to some degree in all the Balkan languages, with Turkish as the likely conduit for the introduction of this *m*-reduplication pattern into the Balkans. For the Balkans it has been studied most thoroughly in Bulgarian (Grannes 1996: 259–286), though there are very detailed studies too for some Greek dialects, especially the Bythinian Greek in northwestern Asia Minor (Konstantinidou 2004), and Stolz 2006 discusses it for the Balkans, with particular attention to its Turkish origins as far as the Balkans are concerned.

Examples abound in the Balkans; a sampling, with an indication of sources where possible, is given in (4.14), where some of the translations give a feel for the colorful, expressive, colloquial, and somewhat pejorative, character that these formations can have – in some instances the formations reflect language-particular embellishments of the basic *m*-doubling pattern, either with added words or with univerbation into a single unit, or the like:

(4.14)

| | |
|---|---|
| Alb | shiri-miri 'confusion' (Schuchardt 1888: 68; Meyer 1891: s.v) |
| | cingra-mingra 'trivia' |
| | çikla-mikla 'tiny bits and pieces; crumbs; trivia' |
| Blg | knigi-migi 'books and such' |
| | skandal-mandal 'scandals and stuff' (Grannes 1996: 278) |
| | snjag-mnjag 'snow and such' (Grannes 1996: 278) |
| Mac | OBSE-mOBSE 'the OSCE and the like' (Prizma 2015)[258] |
| | knigi-migi 'books and such' |
| Aro | sare-mare 'salt and such' (Capidan 1932: 524) |
| | carne-marne 'meat and such' (Capidan 1932: 524) |
| Rmn | ciri-miri 'confusion' (Meyer 1891: 406) |
| Rmi | bajraktari-majraktari 'standard bearers and other such people' |
| | pajtoni majtoni 'carriages and such' |
| | sluge-mluge 'all kinds of servants' (Cech et al. 2009: 216, 232) |

257 The occurrence in Basque, however, is problematic.

258 OSCE is the Organization for Security and Co-operation in Europe (Macedonian *Organizacija za Bezbednost i Sorabotka vo Evropa*), an inter-governmental organization. The speaker, a corrupt official, was mocking the election observers sent by such organizations.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Jud livro mivro ‘books and such’ (Varol Bornes 1996)
sapatos mapatos ‘shoes, shmoes’ (Bunis 1983: 121)
Grk τζάντζαλα μάντζαλα ‘rags and such; useless stuff’
τα σάνταλα κι τα μάνταλα ‘stuff and things’ (OEGrk, Ronzevalle 1911: 441, 1912: 156; with definite article τα and κι ‘and’ (StdGrk και))
σούρδου μούρδου ‘topsy-turvy’ (Levkas dialect; Meyer 1891)
σούρδου (μ)πούρδου ‘topsy-turvy’ (Chios dialect; Meyer 1891)[259]
η σάρα και η μάρα ‘Tom, Dick, and Harry; ragtail and bobtail’ (Meyer 1891; with definite article η and και ‘and’)
άρα μάρα ‘who cares?’
άρες μάρες (κουκουνάρες) ‘nonsense’[260]

Some of these have direct parallels in Turkish, so that not only the pattern but also some of the pieces in particular formations may derive from Turkish elements; this is likely the case with *shiri-miri* and *ciri-miri*, since Meyer 1891: 406 cites Turkish *şur mur* with the same meaning. Turkish *şur* means, among other things, ‘tumult, uproar, commotion’ (Redhouse 1968: s.v.; Ayverdi & Topaloğlu 2006: s.v.; Akalın & Toparlı 2005: s.v.), so this seems to be a straightforward reduplication with *m*-. The Greek σούρδου (μ)πούρδου appears to be at least influenced by if not based on the Turkish.

Some of the languages have a considerable degree of productivity for this construction.[261] Thus, for example, Demetrius Vyzantios, in his 1836 Greek play *Η Βαβυλωνία* (‘Babylonia’), a work that has dialect-based miscomprehension as a recurring theme, uses the *m*-doubling construction quite frequently and for particular effect with a variety of base words;[262] one finds in it, for instance, καφέ μαφέ ‘coffee and such’ and πιπέρι μιπέρι ‘pepper and such,’ based on the nouns καφές ‘coffee’ and πιπέρι ‘pepper,’ respectively, ἔγνωκας μέγνωκας ‘what’s with this ἔγνωκας ?!’ (in mocking a character using Ancient Greek forms), based on a verb,[263] and σουκράτη μουκράτη κυδίδη μυδίδη ‘what’s with Isocrates (and)

259 This form is still current, according to Aleka Akoyunoglou, a Chiote speaker. Note that we include this under *m*-reduplication since Greek μπ ([b]) for many speakers now, and probably many more in earlier times, can have a nasal onset (see §5.4.4.1).

260 The additional word here, κουκουνάρες, means ‘pine cones; pine nuts’ and was added surely just for the rhyme; see Joseph 1985a for discussion of other proposed etymologies for άρα μάρα and άρες μάρες.

261 The Albanian word *xhuxhmaxhuxh*, with the meaning of ‘very short old man [in folklore] with a long beard who lives underground; dwarf,’ looks like an *m*-reduplication-like coinage based on *xhuxh* ‘dwarf,’ a Turkish loanword (StTrk *cüce*). Another possibility is that it is based on Turkish *Ye’cuc ve Me’cuc* ‘Yecuc and Mecuc,’ the Biblical ‘Gog and Magog’ but also representing the ‘name of a dwarf people supposed to appear at the Day of Judgment’ (Redhouse 1968: s.v.). So any connection with the ‘dwarf” word is secondary, presumably influenced in some way by *xhuxh* for ‘dwarf.’

262 Beaton 2004: 94, for instance, says “We know from … Βαβυλωνία … that the differences between regional dialects at the time could cause considerable misunderstandings, which are exploited in the play to comic effect.” See also Levy 1980, where some of the particular examples cited here are mentioned.

263 Ἔγνωκας is an Ancient Greek 2SG perfect tense of γιγνώσκω ‘know,’ a form that ceased to be current from Medieval Greek onward.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Thucydides?!' (mocking a character reading Ancient Greek authors), based on truncated proper names. And, in Bythinian Greek, due to the intensive contact between Turkish speakers and Greek speakers and the bilingualism there, there are literally dozens and dozens of attested instances of this construction (Konstantinidou 2004). Although in modern-day Greek the construction is less frequent than in Bithynian Greek, it is still quite productive, as observed by Kallergi & Konstantinidou 2018: 102–121. Moreover, it is extremely productive in Macedonian, as attested to in *Vistinata za Makedonija* (Prizma 2015; Friedman 2019a), where there are at least a dozen such examples, including the *OBSE-mOBSE* example in (4.14). As with other cases like this (see §4.3.7.2.1, regarding intensive prefixation), it is more likely that the borrowed lexical material was the basis for the emergence of a process in the borrowing languages, rather than that the process itself was borrowed from Turkish.

### 4.3.7.3 Other Reduplications

All of the languages show other reduplications with various forms and various meanings. Romanian, for instance, has *talmeş-balmeş* for 'jumble,' and Greek has (somewhat onomatopoetic) τσαφ-τσουφ 'in an instant.' However, none of these have any systematic status across the several languages nor a specific or demonstrable contact dimension. There is, however, one reduplicative phrase with expressive value that occurs widely throughout the Balkans that shows several variant forms with varied meanings but most centering on inadequate verbal skills. There is some dispute as to the ultimate source, but its widespread manifestation makes it noteworthy in this regard, no matter what its origin is.[264] Relevant forms are given in (4.15):[265]

(4.15) Trk çatra-patra 'incorrectly and brokenly (speaking a foreign language)'
çıtır pıtır 'with a sweet babble (said of the talking of a child), prattling'
çat pat 'a little, some (ability in speaking a language)'
Grk τšὰτ πὰτ (OEGrk, Ronzevalle 1911: 287, 1912: 70)
tšatır patır (OEGrk, Ronzevalle 1911: 288, 1912: 70; also BDJ field notes (Alexandroupolis region, from a twenty-year-old speaker, 1981)
τσάτρα πάτρα 'stumblingly (with reference to speaking a language)'
tšátara pátara (OEGrk, Ronzevalle 1911: 287, "*plus expressif que τšὰτ πὰτ pour dire mal parler une langue*" ('more expressive than τšὰτ πὰτ for saying 'to speak a language badly')

264 Andriotis 1983: s.v., as also Babiniotis 2010: s.v., says it is from a Byzantine Greek σάταλα πάταλα, a phrase referring to artless speech; they see the other Balkan forms as coming ultimately from this. Redhouse 1968: s.v. cites *çatra patra* as being from Greek, but Tietze 2002: s.v. cites *çat pat* 'idem' as being from reduplication with a labial of Old Turkish *çat* 'crack (the sound made when something breaks from being struck [cf. *broken English*]).' Tietze 2002: s.v., and Clauson 1972: s.v. make it clear that *çat-pat* is of Turkish origin. Thus, at the very least, Turkish was involved in the diffusion of these forms, regardless of the possible relation of the Turkish to the Byzantine Greek.

265 There may well be an onomatopoetic, or at least iconic, basis to this formation, with the imperfect doubling representing imperfect command of a language.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

| | |
|---|---|
| Blg | čatăr-patăr '*idiom* (dial[ectal]) so-so, passably, poorly' (Grannes et al. 2002) |
| | čatara-patara 'idem' |
| | čatra-patra 'idem' |
| | čat-pat 'idem' |
| Mac | čat-pat 'idem' |
| Aro | ceat-pat 'so-so, comme si comme ça' (Papahagi 1974: 432) |
| Rmn | ceat-pat 'idem' |
| Alb | çatrapil 'confusion, disorder'[266] |

This pattern may in some sense be related to the *m*-doubling since it involves whole-word repetition of a word with a labial initial, though in this case a *p*-; note too that labials are involved in the intensive prefixation reduplication as prefix-final elements. Still, this particular case is not the borrowing of any sort of pattern per se but is rather a lexical borrowing, albeit interesting and relevant as ERIC loans given the phrase's colloquial character.

#### 4.3.7.4 Conclusions Regarding Reduplication

By its very nature, as argued in §4.3.7.1, reduplication lends itself well to expressive functions, and thus to conversational uses. All of the Balkan evidence cited here illustrates this point well. Given that there are several instances of the borrowing of reduplicative forms in the Balkans, there can be no doubt that reduplication can thrive in situations of language contact, and the positive suggestions of reduplicative patterns spreading make that all the clearer. In these ways, then, reduplication in the Balkans fits in well with the ERIC-loan typology advanced here, both as to its function and as to its role in discourse and conversational interaction. As such, it provides support for viewing the nature of language contact in the region as both intense and intimate.

### 4.3.8 Diminutives, Hypocoristics, and Endearing Terms of Address

Means of addressing interlocutors in an endearing way, and other ways of showing intimacy, as well as respect (in some sense its counterpart), towards a discourse participant, including the use of diminutives, all fall quite naturally into the realm of conversationally based words, inasmuch as they are tied to face-to-face acts of communication between speakers where more than just the exchange of information is involved. They thus quintessentially represent "human-oriented" instead of "object-oriented" interaction. Significantly, and as expected under the ERIC-loan classification of Balkan lexical borrowing advocated here, forms of this sort show considerable diffusion within the Balkans.[267]

266 We assume this derives in some way from Turkish, given that the first two syllables match the Turkish form and they are followed by a *-p-*.

267 For a study of how various standard languages of the Balkans render German diminutive reference in translations of the works of Thomas Mann, see Keller 2010; as interesting as it is, we consider

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Such words are found frequently in several lexical domains discussed above, which by virtue of their inherent reference and/or the social context in which they typically occur, invite the use of endearing address. For instance, they are common in kin terms, as the status of close kin lends itself well to terms of endearment or to notice of relative age or size; see §4.3.1.1 (on 'mother, grandmother'), §4.3.1.2 (on 'brother'), and §4.3.1.4 (on 'uncle') for some examples. Similarly, vocatives also typically occur in social contexts where hypocorism is to be expected, as seen in the examples in §4.3.5. And, expressive and familiar address, seen in §4.3.4.2.2, also is conducive to the use of diminutives and hypocoristics.

In many cases, the usage involves transfer of a foreign item from the typically loving context of the nursery or home or playful interactions to broader reference outside of that domain, but still invoking the original intimacy. Derivatives of Turkish *canım* 'my soul,' *oğlum* 'my son,' and Albanian *bir* 'son' are discussed in §4.3.4.2.2. In many instances, Turkish is the source of other words of this sort; for instance, Bulgarian dialects (Grannes et al. 2002: 15) show Turkish *ata* 'father' used as an 'intimate and respectful term of address for a man, father.'[268] Greek is the source of *kukla* 'doll' and the form with a possessive pronoun, *kuklam* 'my doll' (Greek κούκλα μου, canonically but κούκλα μ in a northern dialect like Constantinople Greek – but also identical with the 1SG.POSS suffix of Turkish), in Judezmo of Istanbul (Varol Bornes 2008: 393), though borrowed as well into Turkish and thus possibly not directly from Greek. Cf. also Romani *Devlam* 'O my God,' cited in §4.3.3.1.1.

Related to this is the use of a range of politeness formulas of Turkish origin throughout the Balkans,[269] as noted by Skok 1935: 254–255 and Mirčev 1963: 76, and as discussed for Bulgarian by Grannes 1969. Grannes observes that the forms *efendi* 'sir' (Trk *efendi*), *efendim* 'my sir, milord' (Trk *efendim*), *gečmiš ola*! 'a wish expressed to persons who have been ill or been through some calamity; get well soon; a speedy recovery' (Trk *geçmiş ola!*, lit., 'passed may.it.be' [Modern Trk *geçmiş olsun*]), *kuzum* 'my dear' (Trk *kuzum* 'my lamb!'), and *šerif aga* 'noble lord' (Trk *şerif ağa*) were all widely used colloquially in Ottoman times in Bulgarian, but also, we can add, in all the Balkan languages.[270] Moreover, Balkan languages such as Greek, Macedonian, and Albanian have all calqued *geçmiş ola/olsun* in various ways: Grk περαστικά (lit., 'passingly'), Mac *da ti pomine* 'may [it] pass you' and Alb [*qoftë*] *të kaluara* '[may it be] passed,' all used to wish someone to get well.

Relevant here too is BSl and Alb *temane*, the word for an 'oriental salute made with a bow and touching the fingers of the right hand to the lips and then to the

this to be more a contribution to translation studies than something directly relevant to Balkan linguistics as conceived of here.

268 The sharing of terms of endearment suggests a social milieu in which adult speakers, most likely women, of one language interacted with children of another language in such a way that other adults, again most likely women caretakers of these children, could pick up these terms for use in their own language. See §4.3.10.1.2.2 for another case pointing to such a social scenario.

269 See also §4.3.4.2.2 and §4.3.10.1.2.2 on shared greetings across the languages.

270 As noted above (§4.2.1.6), while the Greek form, αφέντης is of Turkish origin in this form, the Turkish source word is itself based on Greek αὐθέντης (in its Middle Greek form, pronounced [afθendis]), which in Ancient Greek meant 'perpetrator, murderer; absolute master, ruler' and later just 'master, ruler'; thus αφέντης is a reborrowing, from Greek into Turkish and back into Greek at a later date.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

forehead,' ultimately from Arabic but in the Balkans by way of Turkish. This gesture of respect and politeness was known in Ottoman times and remains to some extent in post-late-Ottoman Rumeli (Albania, Bulgaria, Kosovo, Macedonia, and Thrace), where it is part of many Muslim wedding ceremonies where the bride shows respect to the groom. It also figures in the meeting of Bai Ganyo and the Czech historian Konstantin Jireček (Konstantinov 2010).[271]

Finally, one key way in which diminutivity is expressed in the Balkans is through suffixation, and there are many instances of such suffixes diffusing across the languages. Several of these are quite clear-cut but one is particularly notable – and controversial.

To cover the straightforward cases first, Matras 2009: 210 observes that "Romani dialects borrow a series of agentive and diminutive affixes from various contact languages," although the only example he gives is from Central European Romani (Sinti) and involves a Slavic feminizing suffix: *Sint-ica* 'a Sinti woman.'[272] Slavic is the source of several diminutive suffixes in Romanian, as documented by Puşcariu 1902, such as *-işcă* (as in *morişcă* 'coffee-mill'; for the suffix, cf. Russ *voriška*, based on *vor* 'thief'), and in the case of the diminutive suffixes in *-Vc* (*-uc*, *-oc*, etc.) where native Latinate material is involved etymologically, Puşcariu (p. 142) argues that "the influence of neighboring languages cannot be denied [and] through Slavic influence the Romanian *c*-suffixes acquired greater vitality."[273] Further, the South Slavic diminutives *-če* and *-ko* occur in Albanian as "suffixes of endearment" (Newmark et al. 1982: 172) in borrowings such as *dajko* 'uncle,' but also in native Albanian words such as *vëllako* 'brother,' *birçe* 'sonny boy' and *nipçe* 'nephew.' These same suffixes also occur in Judezmo, e.g., *Avram* > *Avramche, Binyamin* > *Benko* (Bunis 2003: 224–225). Here it is worth noting that the Turkish diminutive *-çe,* which is invariant (e.g., *divançe* 'a small collection of poetry'), is of Persian origin, while the South (mostly Balkan) Slavic could be native or influenced by the Turkish (cf. Sawicka 2021).

Albanian is not just a recipient language as far as diminutives are concerned. Megara Greek, as noted above in §4.3, has borrowed the Albanian diminutive suffix *-zë* in, e.g., λιγάζα 'a little,' and Kyriazis 2012b notes that *-zë* occurs more widely in other Greek dialects in close contact with Arvanitika (though some instances, e.g., βάιζα 'girl' (Alb *vajzë*) simply have the suffix integrated into a borrowed Albanian word).[274]

The Turkish nominal diminutive *-CIK* is the basis for the Greek somewhat slangy and affective nominal/adjectival suffix -τζικος that occurs in λαουτζίκος

271 The word occurs in various forms, e.g., *temana*, *temene*, *temane*, *temenna*; see Grannes et al. 2002: s.v. In the novel *Bai Ganyo*, set in late nineteenth-century Europe, Konstantinov has its eponymous hero greet the famous Czech historian with a *temane*. In so doing, he is transforming a gesture that once signaled urban sophistication into a marker of rural bumpkinhood. On the linguistic versions of this process, see §4.4.1 and references therein.

272 Though see below on the relationship of the feminizing suffix to diminutivity.

273 "*Man darf indessen einen Einfluss von Nachbarsprachen nicht ableugnen . . . Durch slavischen Einfluss bekamen die rum. c-Suff. grössere Lebenskraft*."

274 Vasmer 1941: 122 notes some instances of *-zë* in place names in Attica, e.g., Μπελίζα, from Slavic *Běla* 'white' plus *-zë*, or Καλογρέζα, from καλόγρια 'nun' plus *-zë*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'riff-raff, general populace,' based on the Greek noun λαός 'people,' and possibly μασκαρατζίκος 'young rogue; rascal' alongside μασκαράς 'rascal' (so LKN: s.v. μασκαράς).[275] It generally shows up only in Turkisms, e.g., WRT *kapi* 'door' *kapicik* '[small] back door' gives BSl *kapidžik* (Mac also 'gateway between courtyards'), Alb *kapixhik*, Aro *capiğikje*.[276] However, the form *dadedžik* 'daddy' (Rmi *dade* 'father.VOC' + Turkish diminutive *-cik*) does occur in some dialects of Romani (Cech & Heinschink 1999: 150–153; Friedman 2013b). Jud *kavedžiko* 'a little coffee' makes more sense as the Turkish diminutive + *o* than as Trk *-ci* + Jud *-iko* (*pace* Bunis 1999: 639).[277] Alb *nenexhik* 'mint' (Trk *nâne* 'idem') is probably a local Turkism, borrowed as such into Albanian. Similarly, in Albanian *bërxhik* (Meyer 1891: s.v.) 'the short span between thumb and first finger,' and similar forms, while there is a possible Slavic connection (but cf. Skok 1973: 31), the suffix could have been influenced by Turkish. Cf. also Sawicka 2021 on the possible influence of Turkish *-çe* on the Macedonian diminutive *-če*.

A different source for diminutives is Italic (Latin or Romance) suffix, Latin *-ulla*, found in Greek -ούλα, e.g., ξανθούλα 'blondie,' γατούλα 'kitty.' Greek masculine and neuter forms, -ούλης/-ούλι, respectively, also occur, either built on -ούλα or from Latin gendered forms, e.g., masculine *-ullus*, which is the source of Macedonian (originally from just south of Ohrid, north to Prilep then down just east of Voden (Grk *Édessa*) and Neguš (Grk *Náousa*)) *-ule*, as in *jagnule* 'little lamb' (cf. *jagne* 'lamb'), *detule* 'little child' (cf. *dete* 'child'). This suffix has become productive in Macedonian informal speech, e.g., *kafule* 'small coffee-house.' Albanian *zheg* 'heat wave, oppressive heat' (from Slavic), has a regional variant *zhugull*, presumably with this suffix.[278] While the Greek form is most likely directly from (Late) Latin, the Macedonian is most likely from Balkan Romance (Koneski et al. 1968: 422–423, 538). The Greek form also entered Judezmo, as in the woman's name *Rikula* < *Rika* < *Rivka* (Bunis 1999: 82).

We can also note that the diminutive suffix *-ache* in Romanian is a borrowing from the Greek suffix (neuter) -άκι, (masculine) -άκης (Puşcariu 1902: 223). This

275 As an alternative etymology for -τζικος in this word at least, Andriotis 1983: s.v. suggests a derivation within Greek from the Turkish personal/occupational suffix *-CI* (see §4.2.2.4) with the Greek adjectival suffix -ικο-, despite the apparent absence (to judge from the evidence of various large Greek dictionaries, especially Babiniotis 1998, LKN, and Charalambakis 2014), of a Greek *μασκαρατζής. Even though μασκαράς is said (Andriotis 1983: s.v.; Babiniotis 1998: s.v.) to be directly from Venetian *mascara* 'masked person,' not Venetian via Turkish, other sources, including LKN 1999: s.v. and Babiniotis 2010: s.v., recognize two distinct words with the form *maskara-* in Greek: the Venetian-derived word for a masked person, and an Arabic-derived word, via Turkish, meaning 'buffoon; rascal' (Trk *maskara*, Arbc *mesḫare*). This is consistent with the situation in Turkish, where *maskara* from both Arabic and from Venetian, in the respective meanings, is attested (Ayverdi & Topaloğlu 2006: s.v.). Given that *-CIK* is a productive diminutive suffix in Turkish, it is possible that a form *maskaracık* was borrowed, at which point the adjectival -τζικος could be an adaptation from a nominal suffix -τζικο.

276 In the Ottoman Balkans *kapidžik* was a way for neighbors (especially women) to visit without having to go out into the street, and it was also a means of escape in times of trouble.

277 So also regarding Greek -τζικος; see footnote 275 above.

278 Weigand 1926 gives examples of all sorts of realizations of *-ul* in the Balkans, but many of these are not connected with the Italic diminutive suffix.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

suffix also occurs in Judezmo, e.g., *Avram* > *Avramaki* (Bunis 2003). Bunis 2003: 222 also suggests the possibility that the Judezmo diminutive *-achi,* e.g., *haham* 'scholar' > *hahamachi* 'quite a learned scholar,' may be from the Spanish augmentative *-acho*, influenced by Grk -άκι, Trk *-cik*, and/or Trk *-çe* and BSl *-če.*

The suffixes Alb *-ush-*, BRo *-uş-*, Slv *-uš-*, and Trk *-Iş-* are all involved in various forms of derivations that involve diminutives and expressives, e.g., Alb *zonjushë* 'miss, mademoiselle' (from *zonjë* 'lady, Mrs., mistress [of the house]'), Mac *liguš* 'snot-nosed kid' (from *liga* 'slime, snot'), Rmn *cățeluş* 'doggie' (from *cățel* 'puppy'). In the case of Turkish, the suffix involves high vowel + *ş*, ordinarily subject to vowel harmony. However, in WRT, precisely *-uş-* surfaces as an invariant, e.g., from *Aluş* (hypocristic for *Ali,* vs. StTrk *Aliş*). Directionality versus heritage is difficult to establish except for the probably Balkan influence on the WRT form.

A diminutive suffix that has diffused quite widely, like the Turkish ones discussed in §4.2.2.4, but best treated here due to its function, is the one that is phonetically [-it$^{s}$a], found in all of the Balkan languages. While on the one hand its specific form, with a dental affricate, aligns it with the expressive phonology discussed in §5.7, its relevance here lies in its function and its diffusion throughout the various languages. Examples of this suffix (sometimes with other diminutive suffixes as well) from a range of languages include the items in Table 4.18.

As these examples indicate, these forms with [-it$^{s}$a] show typical diminutive functions, such as marking smallness in size, smallness in age (i.e., youth), low social status, endearment, and the like.[279]

In terms of origins, it is generally held, uncontroversially, that as far as Balkan Slavic is concerned, the suffix is the regular outcome, via the third (progressive) Slavic palatalization, of a feminine suffix with the form **-īkā*.[280] This suffix is

Table 4.18 *Examples of [-it$^{s}$a] in the Balkans*

| | |
|---|---|
| Alb | rrugicë 'alley' < *rrugë* 'road' |
| Blg | ribčica 'little fish' < *ribka* 'small fish' < *riba* 'fish' |
| Mac | rečica 'small river' < *reka* 'river' |
| Rmn | casuliță 'wee house' < *casulă* 'little house' < *casă* 'house' |
| Megl | kudíță (also: kudičkă) 'little tail' < *koadă* 'tail' |
| Grk | πατατίτσα 'little potato' < πατάτα 'potato' |
| Rmi | harica 'a tiny bit' < *hari* 'a little' |
| Jud | amanitsa 'gragger, noisemaker used at Purim' < *Aman* 'Haman' (the villain in the Book of Esther) |

279 See Jurafsky 1996 on the typical semantic range of diminutives.

280 The *-ī-* in this suffix could derive from a diphthong, thus from **-eikā*, or from a sequence with a laryngeal, thus from**-iH-kā*. This suffix is likely a composite, since there is a PIE adjectival suffix **-iko-*, suggesting a segmentation as **-i-ko-*, and thus **-ei-kā* or **-iH-kā*, at which point the diminutive value of the **-ko-* part, as discussed, e.g., by Jurafsky 1996, becomes relevant.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

found elsewhere in Slavic, mostly as a feminizing suffix (e.g., Russ *tsaritsa* 'tsarina'), but in some instances with diminutive(-like) value, at least in origin, as in Russ *jagoditsa* 'buttock, nipple' (cf. *jagoda* 'berry,' thus lit., 'little berry'), *bessmyslitsa* 'nonsense' (diminutive as dismissive or belittling), or *ptica* 'bird' (an old diminutive, cf. OCS and ORuss *pъta* 'bird'), and its Slavic realization as a diminutive is the likely source of the Albanian and Balkan Romance suffix.

There is, however, and has been for over 100 years, considerable controversy as to the origin of -ίτσα as far as its occurrence in Greek is concerned. This is not the place to rehearse all the details of the scholarly dispute, but emblematic of the controversy is the fact that George Hatzidakis changed his mind several times throughout his career, vacillating between taking -ίτσα as a Slavic borrowing and treating it as a Greek-internal development (Georgacas 1982: 31). Similarly, the most authoritative etymological dictionaries offer mixed results, with Andriotis 1983: s.v. (as also in earlier editions) being convinced that it is a borrowing from Slavic *-ica* while Babiniotis 2010: s.v. takes it as being of Greek origin. It is recognized by all that there are lexical items of Slavic origin in Greek, at least 250 outside of the dialects (Georgacas 1982: 45), though most are not in common use now and are best attested in northern dialects; these include βερβερίτσα 'squirrel' (cf. BSl *ververica*), μουσίτσα 'gnat, midge' (Slavic *mъšica*, diminutive of *muxa* 'fly'), and νουζίτσα 'leather strap, belt' (cf. Srb *uzdica* 'rein, bridle'). There are also many Slavic toponyms in Greece (Vasmer 1941), e.g., *Granítsa*, *Stemnítsa*, and *Tsernítsa*. Still, in his definitive collection of dialect and historical material, Georgacas 1982 argues at great length that apart from the clear loans, Greek diminutive -ίτσα has a Greek source, deriving from a colloquial late Koine (c. fourth century CE) palatalization and affricatization (suggested by Coptic borrowings from Greek, e.g., *sibōtos* from κιβωτός 'ark,' *siθára* from κιθάρα 'lyre,' *epēsi* from ἐποίκιον 'farmstead; hamlet') of the -κ- of the Ancient Greek diminutive suffix -ικιον. The fact that there is a full complement of gendered suffixes in Greek derived from the nucleus -(ι)τσ-, specifically neuters -ιτσι (e.g., κορίτσι 'girl,' cf. κόρη 'girl, daughter') and masculines -ιτσης (e.g., the proper name Θεοφιλίτσης, derived from Θεόφιλος) and -ίτσας (e.g., the proper name Ζαχαρίτσας, derived from Ζαχάριος), alongside the feminine -ίτσα, could suggest an internal Greek origin; however those could also be elaborations from a starting point of Slavic origin.[281] Similarly, Georgacas 1982: 30–31 was persuaded by the widespread, truly pan-Hellenic, distribution of -ίτσα as opposed to the far more localized dialect geography of clear Slavic loans, and by the absence of Slavic loans from several parts of the Greek-speaking world, e.g., southern Italy and the Pontic areas, as opposed to the presence of -ίτσα elsewhere, but one could simply appeal to spread internally within Greek, from dialect to dialect, to explain the distribution, a scenario Georgacas himself even endorses in some instances (p. 31). Moreover,

281 And there are others, such as the extended feminine form -ιτσαινα, and various adjectival suffixes, e.g., -ιτσινος and -ιτσικος. Some seemingly related suffixes may have a foreign origin; for instance, the adjectival -ούτσικος as in καλούτσικος 'good-ish' (cf. καλός 'good') is generally agreed to be from Italian *-uccio*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the Coptic evidence is not as compelling as Georgacas suggests: the forms he cites come from the Sahidic dialect and are spelled with the grapheme called *shima*, a letter that in Sahidic seems to represent a Coptic palatalized velar; these loans, therefore, could simply reflect some degree of fronting, but not anything like affricate value, for a velar in Greek before a front vowel, as found in all of the loans.[282] One is left, therefore, with no compelling evidence for Georgacas's account, although it is accepted in Babiniotis 2010.[283]

Further, it strains credulity to suppose that Greek would have an etymologically unrelated suffix that matches the Slavic-derived one so exactly. So even if there is a plausible Greek source, the Slavic suffix, which was known in Greek, could well have enhanced the adoption of a fronted (and by then possibly affricated) variant of the -ικιον suffix and allowed it to emerge and take hold in its affricated form. The chronology of the first actual appearances of -ίτσα in written materials would accord with such a view, as it is found first in ordinary vocabulary in the twelfth-century poems of Θεόδωρος Πρόδρομος (Πτωχοπρόδρομος), e.g., μικροτερίτζιν 'very small,' and in personal names as early as the ninth century (Βοϊδιτζης, in 838CE (Georgacas 1982: 39)). It should be noted that a Greek-internal source for -ίτσα would further mean that – in principle – at least some instances of *-itsa* in other Balkan languages could have been borrowed from Greek. Nonetheless, diffusion ultimately from a Slavic source remains, for most scholars, the most plausible scenario for the spread of an *-itsa* diminutive suffix throughout the Balkans.[284]

### 4.3.9 Taboo Expressions, Insults, and Other Terms of Abuse

For our purposes here, *taboo* refers to socially negatively sanctioned expressions that are sometimes labeled *obscenities* or *swear words*, as opposed to, for example, the kinds of taboo expressions that refer to dangerous animals, religious practices, etc.[285] Thus, for example, the use among Albanian Muslims of the euphemism *mish miku* 'meat friend.ABL' for *mish derri* 'pork' (lit., 'meat pig.ABL') – in order to

282 The Coptic material, however, fails to explain the change of /k/ to /ts/, which reflects the third palatalization of Late Common Slavic, i.e., precisely the period when Slavic speakers occupied most of what is now Greece.

283 The contrary opinion of Andriotis cited above was formed before Georgacas's study, so it is not unlikely that he would have been persuaded had that work been available in time to be taken note of in the 1983 edition of his etymological dictionary.

284 See Joseph 2015b for a full accounting of the pros and cons of this etymological debate.

285 Henderson 1991: 241 citing Parker 1983 makes the point that taboo language and obscenity involved two different vocabularies in ancient Greece, the former religious, the latter not. See especially Parker 1983: 328–331 for a discussion of *taboo* and Ancient Greek concepts of sacred, polluted, unclean, and forbidden. Henderson 1991: 5 also argues that, unlike the situation in Latin, there was no special category of obscene language in Ancient Greek, i.e., no special term for such language, although, as he amply demonstrates, there were precise Ancient Greek equivalents for a variety of terms considered obscene as we think of it (see also Janse 2014).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

avoid mentioning an unclean animal forbidden by Muslim religious law – is outside the purview of this section. Henderson 1991: 7 provides a useful characterization of obscene vocabulary that be cited here:

> The effect of obscenity is to break through social taboos . . . Thus obscenity is most often used to insult someone . . . to make curses, to add power to comedy, jokes, ridicule, and satire. Its efficacy in all these functions resides in its ability to uncover what is forbidden, and thus to shock, anger, or amuse . . .. Very often the exposure is hostile and serves to degrade the object.

Following the methodology employed by Razvratnikov 1979, 1988–1989, we can identify three broad categories of obscene lexical items: (1) body parts, (2) bodily actions and products (sexual and excretory), and (3) abusive and insulting terms and expressions, many of which tend to be culture bound in one way or another.[286] Taking English as our basis of comparison (owing to the fact that English is the language of this book and not to any inherent qualities of English-language obscenities), the basic relevant lexical items are given in Table 4.19.[287]

Taboo expressions and insults, when they enter one language from another, generally do so as ERIC loans, whereas euphemisms can be colloquial or learnèd. Colloquial euphemisms tend to be metaphors, and in some cases the cross-linguistic identity may be typological rather than areal, i.e., due to an inherent property of the thing described rather than language contact. Thus, for example, words meaning 'eggs' can also function in the meaning 'testicles' in various languages (e.g., BSl, Grk, Trk), but this commonality is not necessarily explained as a borrowing, since the shape of the two items in question is suggestive in and of itself (cf. the same in Sanskrit). Learnèd euphemisms, on the other hand, can be either metaphors or borrowings, but do not qualify as ERIC loans, due to their learnèd nature. The difference can itself be a source of humor, as in the following anecdote recounted here in its Macedonian version, where a learnèd borrowing is contrasted to a colloquial expression:

> A young woman from a village is boasting to a girlfriend that she has had an affair with an intellectual. The girlfriend asks what it was like. She explains that the

286 Thus, for example, a term such as Slavic *kurva* 'whore' (which has been borrowed into all the non-Slavic languages of Eastern Europe) presupposes a certain set of cultural values and institutions. Similarly, an epithet meaning 'big-dick' is taken in some languages (e.g., Ancient Greek, cf. Henderson 1991: 76; Dover 1978: 125–135) as insulting, implying the possessor is stupid or like a (dumb) animal, while in others (e.g., modern English *donkey-dick* or Albanian *karderr* 'pig-dick' (lit., 'dick-pig') Tupja 2004), it is taken as an admiring description of virility (vs. Trk *eşeğin siki* 'bullshit!', lit., 'donkey-dick'); cf. also English *son of a bitch* versus the corresponding Chinese *wáng bā dàn* 'son [lit., 'egg'] of a turtle.' According to one explanation, a turtle leaves its eggs after laying them, and so the implication is that the addressee has no legitimate family.

287 We use words in these lists and in our translations that are themselves obscene and slangy in English in order to capture the flavor of the Balkan terms.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.19 *Classes of taboo items in the Balkans*

| | | | | |
|---|---|---|---|---|
| 1 | BODY PARTS | | | |
| | prick/cock[288] | cunt | | |
| | balls | tits | | |
| | ass/arse (incl. asshole) | | | |
| 2A | BODILY ACTIVITY VERBS (all sexual) | | | |
| | fuck | suck/eat | (irrumate)[289] | jack off |
| 2B | BODILY FUNCTIONS AND PRODUCTS (all used as both nouns and verbs in English) | | | |
| | shit | fart | piss | cum |
| 3 | INSULTS (PEOPLE) | | | |
| | bastard – bitch | | | |
| | faggot – dyke[290] | | | |
| | whore – pimp | | | |
| | [stupid – nasty][291] | | | |

> intellectual had a *penis*. The girlfriend, unfamiliar with the term, asks what it means. The young woman replies that it is just like a village *kur* only smaller.[292]

The taboo on mentioning certain human organs connected with sexual activity or excretion (as well as the activities themselves and their products) is not by itself a modern phenomenon in Europe, although the contexts in which these expressions appear have varied over time. Thus, for example, the language of Aristophanes abounds in expressions that even today are forbidden or self-censored in roughly equivalent Euro-Atlantic public media (Henderson 1991).[293] Even in Aristophanes, however, certain lexical items were used precisely for their shock value. At issue, then, was not whether the words were taboo, but the contexts in

288 As is clear from this table, English, unlike all of the Balkan languages, does not have an unambiguous obscene term meaning 'penis.' Unlike *prick*, *cock* does not double as a metaphorical term of abuse in the manner available to *cunt*, as revealed in the difference between "Eat/Suck my. . .!" and "You goddamned/stupid. . .!"

289 There is no special verb in English for the active role in fellatio aside from this Latinate loan. The term is included for the sake of completeness.

290 The pejoration of same-sex relationships is a complex issue that had different instantiations at different times in the history of the Balkans. An account of the terminology is beyond the scope of this section, but the English-language equivalents are included in the schema for the sake of basic completeness.

291 Unlike, e.g., sexual promiscuity, which in some cultures may be valorized for one gender and not for another, or may (like sexual orientation) simply be irrelevant, the qualities of stupidity and nastiness are inherently negative. Consider in this light the Macedonian expression *em gol em zol*, lit., 'and naked and evil,' i.e., 'stupid and nasty.'

292 The anecdote encapsulates a number of assumptions about the opposition rural/urban, masculinity, intelligence, etc., but our point here is that although *kur* happens to be native, an ERIC loan could function in this position, but would not make sense in place of *penis*.

293 In the United States, television and radio broadcasters generally avoid the words *cocksucker, cunt, fuck, motherfucker, piss, shit*, and *tits* – the so-called "Seven Dirty Words"; US comedian Lenny Bruce was arrested for using these words (plus *balls*) in his act in 1966. We can note in passing that *turd*, however, is apparently not subject to such restrictions.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

which they could be used for certain kinds of dramatic effect. To this can be added the fact that in many languages, the native terms for so-called private parts are simply the noneuphemistic ones, but they become obscene either in context or as a result of contact with another culture's prudery (cf. the removal of most of Japan's Shinto phallic fertility shrines during the Meiji period as a result of European influence; see Kinoshita & Palevsky 1992: 30). That said, however, it is also certainly the case that sexual and excretory organs and their activities form a special category in all the European (and many other) languages, and they are deployed in various forms of verbal abuse in many diverse cultures.

In the context of Balkan linguistics, it is striking that in fact relatively few obscenities are borrowed and in most languages the terms are of native origin, albeit that these terms often go through a cycle that can be identified as a kind of pollution, i.e., a word starts out as a euphemism and eventually becomes so closely associated with the obscenity that it originally stood for that the old euphemism becomes a new obscenity. Thus, for example, Ancient Greek οἴφειν 'to mount,' which continues the Indo-European root **yebh-* 'to have sexual intercourse' (cf. Sanskrit *yabhati* 'copulates,' Slavic *(j)eb-* 'fuck,' Sogdian *a-yāmb* 'commit adultery'), was already considered to be a Doric provincialism by the time of Aristophanes, when βινεῖν was the *vox propria* for 'fuck' (Henderson 1991: 35).[294] In Modern Greek, however, that word is completely obsolete, and the verb γαμώ, originally 'marry,' is now the verb for 'fuck.'[295] In Balkan Slavic, however, as in most of the rest of Slavic, most of the languages have preserved *(j)eb-* as the obscene verb. In the sections that follow, we give coverage for both native and shared vocabulary precisely because the semantic field as a whole is typically ERIC, but this particular semantic field is quite specific, is not covered in any other Balkan handbook, and is subject to folk beliefs such as the claim that many of them are of Turkish origin, e.g., Grannes 1969, who notes (quoting from 1996: 109) that "Skok 1935: 254–255 *affirme que dans les langues balkaniques les jurons sont souvent d'origine orientale*" ('Skok affirms that in the Balkan languages, swear words are often of eastern origin'). This is not so much a statement of fact as of sociocultural attitude. While certain Turkish obscenities have penetrated into the Balkan languages, often with altered meaning, the actual patterns of borrowing are in fact much more complex. Given the richness of the field, however, we only cover some basic terminology.

### 4.3.9.1 Body Parts

Balkan languages show both differences and similarities in the distribution of dedicated versus metaphorical terms for body parts. Thus, for example, both English and

294 It has been proposed that the etymology of βινεῖν is related to beating or violence, but the exact details of its origin are unclear. IE **yebh-* orginally meant 'enter, penetrate,' a meaning preserved in Tocharian (TochA *yow-*, TochB *yāp-* 'to enter, set [of sun]'); cf. also Luv *ipatarma* 'west' (Mallory & Adams 1997: 508).

295 This usage occurs in late antiquity (Lucian), and was established by Byzantine times (Buck 1949: 279).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the Balkan languages have specific terms for vagina/vulva – English *cunt*, Slavic *pizda* and its derivatives (*pička* etc., Bulg *putka* is of different origin, BER VI), which is cognate with Albanian *pidh* (Hamp 1968; but note also *piçkë*, *piç,* from Slavic), Grk μουνί (likely from Vtn *mona*), Trk *am*, Rmi *mindž*, BRo *pizdă/kizdă* (a borrowing from Slavic). Romani is the source of BSl *mindža*, which in turn is the likely source of Trk *mınca/minco* (since final *-inç* – devoicing is automatic – is permissible in Turkish). On the other hand, terms for 'penis' in English such as *cock, dick, prick* all have other potential meanings (although in context they are unambiguous), while in the Balkan languages SSl *kur* and its derivatives, albeit etymologically related to CoSl **kurŭ* 'cock,' are unambiguous (but see Loma 2004 for alternative explanations). The phallic qualities attributed to the rooster are such that the parallel between English and Balkan Slavic is merely typological. Bulgarian also has East Slavic *huj*.[296] The bird metaphor is also the source of Balkan Romance *pulă* < Lat *pulla* 'birdie, etc.' According to Meyer 1891: 176, Albanian *kar,* which is the *vox propria*, is borrowed from Romani, for which this word preserves the original meaning, cf. Prkt *kāṭa* 'penis.' The word also occurs in Romanian slang *car* 'idem' (Leschber 1995: 158, cf. also *a carici* 'to fuck,' Armjanov 2001: 69, *karam* 'fuck v.' archaic).[297] Of a different source are AGrk πέος, πόσθη, which are derivatives of PIE **pes-* (**pes-os* and **pos-dhā* respectively; so also Lat *penis* < **pes-ni-*), and Rmn/Aro *puță* (which is diminutive) from VLat *pūtium* (cf. Lat *praeputim*), the source also of ModGrk πούτσος (not diminutive); cf. also Alb *pucarak* 'spirited, brave person.' Turkish *sik* is both the noun and the verbal root meaning 'fuck' (cf. *to bone* in English, from *boner* 'erection', and *to ball* from *balls* 'testicles', for typological parallels). For 'testicles/balls,' BSl *made/măde*, Rmn *coaie*, Aro *coalje*, *hărhăndeale, boashe,* Alb *herdhe,* ModGrk αρχίδια (these latter two from PIE **Hórǵhis* 'idem'), Rmi *pele* (Skt *pela*), Trk *taşak* (also = Jud) are unambiguous (but cf. Trk *taş* 'stone'), whereas English *balls, nuts,* etc. are metaphorical (cf. 'eggs' mentioned above). Words for 'arse' are generally native: Mac *gaz*/Blg *găz*, Alb *bythë*, Rmi *bul*, Trk *göt*, Grk κώλος, but note Rmn *cur* (related to κώλος, with rhotacism, or possibly from Latin *colon*), Jud *kulo* (possibly from Greek, though it could well be inherited from Latin; cf. French *cul*).[298] Romanian *găoază* 'ass(hole)' does not have a simple etymology, but at least the influence of BSl seems likely. Various equivalents for 'tits' do not present any specifically Balkan or contact features.

### 4.3.9.2 Bodily Activities, Functions, and Products

In general, like body parts, obscene expressions relating to bodily activities, functions, and products are of native origin in the Balkans. Excretory functions such as 'piss' and 'fart' lend themselves to onomatopoeia, and yet Mac/Blg *prdi/*

296 If Vasmer 1986–1987: s.v. is correct in connecting *huj* with *hvoj* 'pine needle,' then the semantics are like English *prick*.

297 Bulgarian *kara* 'drive forward, etc.' is also a possible source for the Bulgarian obscenity.

298 For 'buttocks,' British English has the unambiguous Germanic *arse* but the North American English *ass* (ultimately from Lat *asinus*) ends up being homophonous with the animal, which in its turn can be a term of abuse.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

*părdja*, Alb *pjerdh*, Grk πέρδομαι (MedGrk denominal πορδίζω), Jud *pedar* are all cognate with English *fart*, thus from PIE **perd-*, possibly itself onomatopoetic. Rmn/Aro *băşi*/*bes* (Lat *vissīre*) is likewise onomatopoetic, while Rmi *kha[n]jarel* (*khand* 'stink' < OInd *gandha-* 'odor'), and Trk *osur-* (cf. **ASR-* (Clauson 1972: 250)) are connected to 'odor.' Rmn *pîrdalnic* (WMuntenia *purdalnic*) 'damned, diabolical, devil' is a borrowing from BSl. The noun is generally derived from the verbal root.

The forms Rmn/Aro *pişa/kishat*, and Jud *pišar* are imitative from VLat **pissiāre*, and Buck 1949: 273 cites the Romance as the origin of SSl *piša-*. The VLat form, through OFrn *pisser,* also gives English *piss*, Russ *písat'* (3PL *písajut*); OCS has *sьcati* (PIE **seik$^{w}$-* 'pour'), which only survives in North Slavic. Alb *pshurrë* is probably also imitative (although Meyer 1891: 420 connects it with IE **sū-ro* 'salty, sour, bitter'; cf. OCS *syrъ* 'cheese,' cognate with Albanian *hirrë* 'whey' (Vasmer 1986–1987: s.v.)). Grk κατουρώ (AGrk οὐρέω) is cognate with Lat *urina*, the source of Eng *urine* via a borrowing. Rmi *mut[a]rel* (= Skt *mūtra-)* < PIE **meu-*, cf. OCS *myti* 'wash' and Eng *mud*. Trk *işe-* also looks onomatopoetic (but cf. Clauson 1972: 255).

Like 'piss' and 'fart,' 'shit' is generally native. The imitative/nursery form *kaka-* is Indo-European (PIE **kakka-* (Pokorny 1959: s.v.), **kak(k)eh$_a$-* (in the terms of Mallory & Adams 1997: 187)) and survives in Romance, Slavic, Greek, Albanian, Celtic, and Armenian, albeit in registers varying from the nursery to the obscene; Rmn/Aro *căca* (V)*, căcat* (N)*/cac, cãcat* are in this latter category, as is Jud *kagar*. BRo *cacat* is the relevant noun, but most Balkan languages have distinct nouns and verbs. For verbs, Alb *dhjes* and Grk χέζω are cognate with one another and also possibly cognate with Rmi *xiel, xlijel, xinel, xendel* (ptcp. *x(l)endo*), if this is related to Skt *had-* 'excrement' (IE **ǵhed-*, cf. Mallory & Adams 1997: 187). Boretzky & Igla 1994: 116 suggest the influence of Grk χύνω 'pour; ejaculate' or χέζω 'shit' on the Romani form (cf. also Paspati 1870: 315). Strandža Bulgarian *nasihesva* 'to make number two [of a dog]' is probably from or at least influenced by the Greek (cf. BER IV: s.v.). BSl *sere*/*sra* 'shit.3SG.PRS/shit.3SG.AOR' is from Common Slavic but is of disputed origin (Hamp 1975a suggests a possible Old Iranian loan; BER VI: s.v. summarizes the problems). Trk *sıç-* is old and native. For the noun, BSl has both Mac *gomno/* Blg *govno* and *lajno*, and the former may ultimately be cognate with Rmi *khul* 'shit' (Skt *gūtha-* (Monier-Williams 1899: s.v.), Prkt *gūh*).[299] ModGrk σκατό (usually plural σκατά; from AGrk σκῶρ, GEN σκατός) has been connected to the Slavic verb cited above (IE **sóḱr*, Mallory & Adams 1997: 186; but cf. Hamp 1975a). Alb *mut* is ultimately from IE **meug/meuk-* 'slip, slide, slime' (Lat *mūcus*). Aro *merdu*, Jud *medra* continue Latin *merda* 'shit' with cognates in Balto-Slavic and

299 According to Hamp 1975a (also cited in BER III: s.v. *lajnó*), the former is ultimately from IE **g$^{w}$u-o-* 'bovine' (cf. English *cow*), while the latter is related to *loj* 'fat.' The original distinction was agricultural: cow manure is useful for fertilizer while human excrement is greasy and unsuitable. Both words were originally adjectives, modifying a word meaning 'excrement, etc.' Murgoski (2013: s.v.) gives more idioms for the *g*-word than for the *l*-word. German *Kot* 'dung' is also cognate with the former.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Germanic (e.g., OCS *smrьděti* 'stink,' English *smart* 'hurt'), but the Romanian continuation is *desmierda* 'caress.' Turkish *bok* (old and native) is also used in Judezmo, which, however, also has Romance *privada* in this meaning. Turkish *bok* has entered Bulgarian in *bokluk* 'garbage, trash' (both literal and figurative), a derivation and meaning that is absent in modern Turkish.

With regard to sexual activities, while all the Balkan languages are capable of expressing various sexual acts, the only act for which all the Balkan languages share an obscene *vox propria* is 'fuck,' and some also have dedicated verbs for 'jack off.' Many verbs and nouns have possible sexual overtones, and there are expressions using nonobscene words that have obscene meanings in the appropriate context, e.g., Greek τσιμπούκι '(Turkish long-stem) pipe' but also 'blow job,' Mac *puši* 'smoke' but also 'give a blow job,' Alb *jargë* 'slime, spit, phlegm, drool' but also 'prostatic fluid (pre-cum), vaginal secretion,' etc. For 'jack off,' Grk μαλακίζομαι, Mac *drka* are dedicated verbs. Rmn *malahie* 'jack off,' from Grk μαλακία, is attested from the seventeenth century, coinciding with the beginnings of the eighteenth century Phanariote domination of the Romanian principalities. As noted above, Slavic *(j)eb-* and ModGrk γαμώ are the relevant words for 'fuck' in Balkan Slavic and Greek respectively. Balkan Romance has *fută/futã* (Lat *futuere*), Turkish has *sik-*, which as a noun is 'cock.' Albanian *qij* (Arv *qienj*) is not from Latin *inclinare* (Meyer 1891: 226) since there is no /l / in the Arvanitika. Orel (1986: 361) suggests Latin *coïre,* but the */nj/* of Arvanitika presents a problem for such a solution. Topalli (2017: s.v.) proposes a native origin from PIE **(s)k(h)ai-* 'hit;' cf. Lat *caedō* 'lie with,' Skt *khidati* 'strike').[300] Balkan Romance, like the rest of Romance, preserves a specifically Latin (or Italic) development, *fut-*. Romani *kurrel* is native, cf. Skt *kuṭṭayati* 'bash, pound,' but note also *del bul* 'hit ass' whence Rmn *a buli*.[301] As noted above, the Turkish root *sik-* is both 'cock' (noun) and 'fuck' (verb). In general, Balkan equivalents of 'cum' (noun or verb) are not dedicated forms, an exception being Rmi *čhorajbe*, a deverbal noun from *čhor-* 'spill,' meaning 'cum' (noun) in both Romani and Macedonian. The (historical) causative *čhoravel* can mean 'ejaculate' or 'urinate,' but the deverbal noun is unambiguous.

### 4.3.9.3 Insults

Unlike the names of body parts, functions, and products, insults are a rich area of ERIC loans, and many of them are indeed from Turkish, or at least were from Turkish into the twentieth century (cf. *pezevenk* 'pimp' and *orospu* 'whore' discussed below). The causative/passive imperative *siktir* which can be translated 'get fucked,' whence also 'fuck off,' is borrowed into all the Balkan languages, although it is merely rude rather than obscene, the source meaning now being

300 With regard to Meyer's proposal, a form **clino* would make more sense than *inclino* in terms of the initial voiceless <q> in Albanian, but the verb is not attested unprefixed in Latin, and, as already noted, there is also the problem of the absence of /l/ from the Arvanitika form.

301 Cf. English *piece of ass*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

unknown in the various borrowing languages.[302] Consider in this context the gradations from *get lost!* to *scram!* to *buzz off!* to *go to the devil/hell* to (British) *piss off!/*(North American) *fuck off!*. The imperative serves as the basis for derived verbs meaning 'to chase off/send someone to hell, etc.' in all the Balkan languages, e.g., BSl *sikter[d]is[uv]a* (with numerous variants), Aro *sictărescu*, Rmn *a sictiri*, Alb *sikteris*, Grk σιχτιρίζω, Jud *siktirear*.[303] The shape of the vowel in the second syllable is itself an indication of oral transmission. In the West Rumelian dialects – in contrast to East Rumelian dialects and the Turkish standard, which is based in part on these latter, as they include Istanbul – high front /i/ is backed to /ɪ/ (realized as schwa in languages with no high back unrounded vowel) in closed syllables. Thus, Albanian has *siktër* as well as standard *sikter*, the Aromanian of Greek Macedonia has *siktãr*, and western Aegean Macedonian dialects have *siktər*. By contrast, Bulgarian and Judezmo have *siktir*, Romanian and the Aromanian of Greece have *sictir*, and Greek has σιχτίρ, all with the East Rumelian vocalism (and Greek with the regular development of κ > χ before a stop). Standard Macedonian and BCMS have *sikter*, as does the Aromanian of North Macedonia. Here the Macedonian /e/ looks like an older schwa that fell together with secondary jer, while the BCMS and Aromanian forms appear to have entered from Macedonian.[304]

Turkish *sik* also occurs in Blg *nasik(i)me* 'I don't give a hoot' from Turkish *sik-im-e* 'penis-my-DAT' plus the Bulgarian directional preposition *na*. BER IV: s.v. *nasikmè* notes that Romani has *me kar-es-te* 'my penis-OBL-LOC,' which, however contains a locative rather than a dative.

Obscenities involving the mother of the addressee are widespread, although the force varies among cultures. Thus, for example in Kilivila, a language of the Trobriand Islands, *kwoy inam* 'fuck.IMPV your.mother' is jocular, *kwoy lumuta* 'fuck your.sister' is serious, and *kwoy um kwava* 'fuck your wife' is a deadly insult (Malinowski 1929: 409).[305] Henderson's observation on the nature of obscenity cited above is especially apt here. Still, the command to go fuck someone or something is widespread or perhaps universal. In some languages, however, the

302 The vowel in the second syllable can show variation, as discussed immediately below.

303 *Sikter* is also the name of a Bosnian punk rock band founded in 1990. BCMS also has the verb *sikterisati*.

304 The Albanian vocalism also might have arrived by such a route, although a specific local Turkish dialectism might also be responsible. Another possibility is that since as an exclamation (rather than a literal imperative), the second vowel in Turkish *siktir* is unstressed and therefore, being also non-initial, is lax and somewhat centralized, this could have contributed to the perception of /e/. Still, the fact that this reflex shows up precisely in Macedonian, Albanian, and BCMS suggests the possibility of transmission from south to north. Initial stress might also explain the schwa in dialectal Albanian and Aegean Macedonian, although there is a WRT tendency to back front <*i*> to <*ɪ*> in closed syllables, which could also be a factor.

305 In Kilivila, the speaker names the addressee's sister, which makes the insult even more serious. In Albanian, using the sister as the object is considered much more insulting than the mother. Here it is instructive to imagine an altercation in English, where, if the speaker were to use 'sister' rather than 'mother' as the DO, it would be less stereotypical, but potentially more inflammatory, since the possibility of its being true would be more imaginable unless the addressee were known to not have any female siblings.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

imperative is at least in competition with an indicative or optative, sometimes involving the first person. Thus, for example, in BCMS, the most common formula involves a first-person singular present plus second person ethical dative – *jebem ti* + DO – while in Albanian, the most normal form is a 1SG OPT (with or without second-person ethical dative) – *[të] qifsha* + DO. In Turkish the optative or gnomic present (*geniş zaman*, see §6.2.4.2.6) are in competition with the 1sg definite (confirmative) past – DO + *siktim* – which is also common. The use of the descendants of the Common Slavic perfect (using what is historically the resultative*l*-participle) in various modern Slavic languages is potentially ambiguous between a past and an optative reading (cf. Friedman 2012b and §4.3.4.1.1 above). Thus the BCMS *jeb'o te pas mater* and the Russ *job tvoju mat'* could both be interpreted as either past resultative or archaic optatives.[306] In terms of Balkan specificities, however, a phrase of the type 'your mother's cunt' – BSl *pička ti majčina, pizda materina, putka ti mamina*, Alb *pidhin e s'ëmës*, BRo *p/kizda mă-tii,* Trk *ananın amı*, Rmi *te dakiri mindž*, Grk της μάνας σου το μουνί – with 'cunt' in the accusative (as the implied object of 'I fuck') in those languages that mark it, is an idiom that, within the European context, is specifically Balkan in its idiomaticity.[307] Imprecations of the type corresponding to English *fuck your mother* involve a first person subject rather than an imperative, but the verb can be optative, preterite, or present depending on the language.

While the phrase meaning 'eat shit' is inherently offensive, in the Balkans it has the idiomatic meaning 'talk nonsense/slander, lie etc.' (cf. Eng *to bullshit*): Mac *jade gomno*, Alb *ha mut*, Rmn *mânca căcat*, Trk *bok yemek,* Rmi *hal khul.* In Judezmo, *komer medra* means 'to suffer in silence, submit to humiliation,' but the expression *medra ke koma* means 'he lied outrageously,' in keeping with the Balkan idiom.

Words meaning 'whore' appear to spread readily. In the Balkans, as elsewhere in eastern Europe, Slavic *kurva* is found in all the languages (as noted in footnote 287; Loma 2004 provides a plausible Greek source for the Slavic, but Slavic is the source for the rest of Eastern Europe). Romani *lubni* generally occurs in slang registers, often with a masculine referent, in which case it means 'faggot' rather than 'whore,' e.g., Trk *lubun*, *lubunya*, Grk λούμπα, λουμπίνα, etc., although in Epirot Grk λούβου preserves the meaning 'whore' (cf. Theodoridis 1966: 133–134). Romanian *bulangiu* 'faggot, bastard/s.o.b.' is from Romani *bul* 'ass' with the productive Turkish agentive suffix (cf. Alb *bythexhi* 'idem' but with the native Albanian base *bythë*). Relations with Venice are seen in the occurrence of Italian *putana* 'whore' in Albanian (*putanë*), Greek (πουτάνα), Aromanian *(putanã*), and

306 Isačenko 1964, who provides evidence that 'dog' is the subject in Russian, and based on the Latin translation of the Hungarian equivalent, notes that the Latin is a subjunctive, i.e., in an optative use, although he neglects to adduce the fact that the *l*-form in the Russian version could simply be an archaic use of the *l*-form as an optative. A similar usage with 'dog' as subject, 'mother' as direct object, and *eb-* as the verb preceded by the DMS is attested in a Wallacho-Bulgarian letter from 1440 (Bogdan 1905:43).

307 'Mother's cunt' figures elsewhere in idiomatic abuse beyond the Balkans, but not in exactly the same idiom, e.g., in Russian it is often the object, explicit or understood, of imperative 'go to.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Macedonian (*putana*). Turkish *orospu* 'whore' and *pezevenk* 'pimp' are treated below.

Mention should be made here of a Balkan term of abuse whose proposed etymologies include all the main branches of the Balkan sprachbund except Indic (Romani), i.e., the ancestor of Albanian, Hellenic, Latin, Slavic, and Turkish. This word, which has a number of meanings, not all of which are pejorative or current, but one of the most common of which is 'bastard' in both the literal and various figurative senses of the word is the following: Alb *kopil*, BCMS *kòpīl* (in Kosovo *kòpilj*), Blg *kópele*, Mac *kopile*, Aro *copil, copelă*, Rmn *cópil* (vs. *copíl* 'child'), Rmi *kopíli*, Trk *kopil*, and also Ukr *kópyl*.[308] The most broad-ranging summary of the various etymologies can be found in BER II: s.v. *kopele i kopile*, but Skok 1972: s.v. *kopīl* should be consulted for additional details. It is not our place here to judge among the various arguments or even adduce them. Our point here is that this is a Balkanism, regardless of its origin, and, moreover, in many instances a shared abusive term.

Turkish was the source of a significant number of insulting terms in all the Balkan languages that were still well attested in the nineteenth century. A few of these, e.g., BSl *budala*, Alb *budalla*, Grk μπουνταλάς 'fool' (< Trk *budala* 'fool[ish]') are still widely understood and employed. Like many other Turkisms, however, most of the terms of abuse have become archaic. Thus, for example, Grannes 1969 cites a number of terms from nineteenth-century Bulgarian, many of which are now archaic, quaint, obsolete, or simply unknown, much like English *rapscallion, guttersnipe*, and *floozy* – all of which are insulting but none of which are current – or *scoundrel*, which is negative, but sounds bookish and does not have the power of its colloquial equivalents. In some cases, however, an abusive Turkism retains its power in some Balkan languages but not others. Thus, for example, Trk *pezevenk* 'pimp' is still very rude in Romani, Albanian (*pizeveng*), and Cypriot Greek, but old-fashioned for most speakers of Balkan Slavic (although some speakers still consider it vulgar), and archaic or forgotten by Balkan Romance speakers and in mainland Greek. Turkish *orospu* 'whore' (BSl *orospija*) is very rude (vulgar) in Albanian and Macdonian, but generally old-fashioned in Bulgarian, mildly abusive in Romani, and archaic or forgotten in Greek and Balkan Romance. Turkish *köpek* 'dog' was still an Aromanian insult as *kjopek* (alongside native *cãne*) in the twentieth century although not recorded in any of the dictionaries (Polenakovikj 2007: s.v.). Romani *džukel* 'dog' is the source of Macedonian slang *džukela* 'street dog, bitch, s.o.b.'

308 Modern Greek is alone among the Balkan languages not having a version of this term in the meaning 'bastard' although, it is among the suggested sources for the word in other languages, e.g., κοπέλι 'child, lad' (BER II: s.v.).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.3.9.4 Ethnophaulisms and Ethnonyms

The term *ethnophaulism* was coined by Roback 1944 in the context of World War Two to mean 'ethnic slur' or 'ethnic term of pejoration.' The line between an ethnophaulism and an ethnonym can be as fine as the difference of context and intonation. As in other parts of Europe, Roms and Jews were objects of prejudice in the Balkans. Thus, for example, the main Balkan languages share *cigan/ţigan/* τσιγγάνος/*çingene* 'Gypsy' (Alb&BSl/BRo/Grk/Trk) and *çifut/čifut/cifut/* τσιφούτης (Alb&Trk/BSl&Jud/BRo/Grk) 'Jew,' which have doubled as ethnonyms and ethnophaulisms. Nonetheless, there are certain Balkan specificities. In the case of τσιγγάνος and related terms, the proposed etymology from Bzyantine Greek ἀθίγγανος 'heretic' (lit., 'untouchable') is problematic on several counts.[309] On the other hand, there are good reasons to judge Turkish *çengene/çingene* as of Central Asian heritage, relating, perhaps, to contact between the early Roms and Turkic tribes, cf. Turkic *čiɣan* 'an old Turkic appellation for low-caste slaves' (Matras 2011: 257). This works well as the source of Greek τσιγγάνος, which is then taken up elsewhere in the Ottoman Empire and beyond. Unmediated *çingene* was also used in the Balkans as a kind of Turkish code-switch in languages other than Greek. Similarly, Turkish *çifut* (ultimately from the Arabism *Yahudî,* from Heb *Y'hudi* 'Judahite, Jew') and its derivatives survive in all the Balkan languages (having entered directly from Turkish, as seen in the word-initial *č-*, whence Grk τσ-), although the meaning 'Jew' survives only in Albanian. In Greek, the word means 'miser,' and modern Greeks are often no more aware of the connection with the original meaning of 'Jew' than are most Americans of the fact that *gyp* 'cheat' began as an ethnic slur on Gypsies (Roms).[310] Romani *džut* 'Jew' is also ultimately from *y'hudi*, but via Iranian (cf. *džuhuro* 'Judeo-Tat'); i.e., the ethnonym entered Romani before the Roms entered the Byzantine Empire.

Terms meaning 'Aromanian' similarly double as ethnophaulisms or ethnic stereotypes. Thus, for example, βλάχος is used in ModGrk to mean 'shepherd' but also 'bumpkin' (perhaps influenced by the ModGrk βλάκας 'idiot' from AGrk βλάξ 'idem'). The Greek Κουτσόβλαχος, literally 'lame Vlah,' is sometimes used for Aromanians, but is generally considered pejorative. Similarly, in BCMS territory, *Cincar* has been used to refer to Aromanians as opposed to Romanians. The origin is said to be the typically Aromanian change of /č/ to /c/ as exemplified by *tsintsi* 'five.' By extension, and not unlike *čifut*, the term is also associated with miserliness, as the Aromanians in these regions were often urban merchants. By contrast, the term *Vlah* referred to Romanian speakers in BCMS territory, but was associated with Aromanian speakers, many of whom were shepherds, in the

309 For instance, the sound change of -θι- ([θι]) to -τσι- ([tsi]) is neither regular nor expected, though the phonosymbolism of Greek τσ ([ts]) discussed in Joseph 1994b (and see §5.7) might provide an avenue for the appearance of the -τσ-, given the societal marginality of Roms and the occurrence of τσ, iconically, as a marginal phoneme in Greek, in some terms for marginalized people. See also footnote 109 above, and Chapter 5, footnote 210.

310 In the southeastern United States, the use of *jew* as a verb to mean 'bargain' is likewise opaque to some speakers.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

southern Balkans.[311] As a result, in modern colloquial Albanian, *vlah* means 'shepherd' and the folk word for Aromanian is *çoban*, a Turkism that in all the other Balkan languages means 'shepherd.' (In the literary language, *Vlah* [pl. *Vleh*] is now used to mean 'Aromanian.') We can note also that Meglenoromanian speakers are the only Balkan Romance group to adopt the exonym *Vlah* (Megl *vla*, pl. *vlaši*) as an autonym, all other groups retaining a word derived from *Romanus* (*Rumîn, Armîn, Rămăn*, etc.). Moreover, in Catholic dialects of BCMS, *Vlah* came to refer to Serbian speakers (as Orthodox Christians, some of whom may have shifted from Romance to Slavic, and some of whom were traditionally shepherds cf. Sikimić & Ašić 2008). In Judezmo, *Blahu* meant simply 'Christian' (Benor 2009; Heb *arel*, literally 'uncircumcised,' was also used; cf. Yiddish and Judeo-Italian *goy* from Hebrew 'nation' – already attested in this meaning in Talmudic times). At present, the use of *Vlah* to refer to a Serb is an ethnophaulism (the corresponding terms for Croat/Catholic and Bosniak/Muslim are *Šokac* and *Balija*, respectively – the former derived from a region in Croatia, the latter from a common Muslim proper name).[312] Albanian speakers did not begin using the term *shqip* and *Shqiptar* until after the Turkish conquest. It is unknown in southern Italy and Greece, where speakers use the term *arbërisht* for their language. Like Greek αρβανίτικα and Turkish *Arnavut*, these forms are all ultimately from *Alban-* (with Tosk rhotacism, a Greek sound change of *l* to *r* – see §5.4.4.9.1 – and Turkish rounding after /v/, as appropriate). Although the form *Šiptar* was normal in Slavic into the 1950s, like other ethnonyms it could be used pejoratively, and in connection with the rejection of Yugoslav attempts to create a *Šiptar* identity in Yugoslavia as distinct from *Albanac* for Albanian of Albania, the term is now exclusively pejorative in Slavic (see Elliott 2017: 145–192). The pejorative Albanian term for Slav is *shka* (pl. *shqe*), which derives ultimately from the Slavic autonym *Slověne* (via the Latin/Greek *Sklaven-*). It is interesting to note that the Arvanitika word for Greek is *shklerisht*, i.e., 'Slavic' (with Tosk rhotacism). There is debate concerning whether this term was brought to the Peloponnese and then reapplied to the new foreign language or whether in fact at the time the term became fixed in Arvanitika, the neighboring foreign language in the Peloponnese was still Slavic. In Greek, Βούλγαρος (Aromanian *vărgăr*) 'Bulgarian' is also used as a slur meaning 'stupid,' and until the nineteenth century, Έλλην(ας), now 'Greek,' meant 'pagan' – a crime punishable by death in Byzantine times – the Middle Greek autonym being Ρωμαίος 'Roman,' now Ρωμιός (Turkish *Rum*), which in the second half of the twentieth century came to have a pejorative sense in Greek meaning 'Balkan bumpkin Greek,' but now has a more positive value, not unlike, perhaps, the ambivalence associated with *cowboy* in the southwest of the United States. In

311 The term *Vlah* as used in pre-Ottoman medieval sources is presumed to refer to Romance speakers. The specifics of that period are complex and need not concern us here (see Fine 1983, 1987: s.v. *Vlach*).

312 We can also note here the term *Morlak* from *mavrov[a]lah* 'black Vlah' – the term *Karavlah* 'black Vlah' also occurs. The Morlaks were Slavic-speaking Orthodox Christians of Romance origin in Dalmatia. The color term in *Karavlah* referred to 'north.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Turkish, the *arnavut* 'Albanian' was stereotyped as stupid, stubborn, and/or violent. Turkish *arnavut inadı* 'Albanian stubbornness' is still a proverbial expression. Any ethnonym can be rendered insulting in Turkish with the addition of *pis* 'filthy' (a Balkan Turkism still in use in Albanian). The Turkish word for 'infidel,' *gâvur* (dialectal *kaur*) could be used as an insult for Christians, but it was also adopted as an autonym by some Balkan Christians.

In Judezmo, a number of nicknames were used cryptoglossically for outsiders, e.g., *los verdes* 'the green ones' for Muslims referring to the fact that the color is sacred in Islam. By extension, the Hebrew-derived nickname *karpasis* 'green vegetables' is also used. A nickname used in Sarajevo was *almesha* (lit., 'plum'), referring to *šljivovica* 'plum brandy,' forbidden by Muslim law but still the most popular alcoholic beverage. An interesting example of fractal recursion (Gal & Irvine 1995; Irvine & Gal 2000; Gal & Irvine 2019) is Judezmo *kweshkos* 'fruit pits, stingy' (from Spanish) to refer to Ashkenazic Jews (cf. *çifut*, etc. cited above). Judezmo words for Christians come from Hebrew, e.g., *arel* 'uncircumcised' (noted above) or *trefán* 'not Kosher' (see Benor 2009 for additional details).

Finally, we can mention the Romani terms *gadžo/gadži/gadže* 'non-Rom (M/F/PL)' and *gomi/gomni* 'non-Rom, peasant.' These terms can be purely descriptive in a neutral context, like the opposition *Yid/Goy* 'Jew/non-Jew' in Yiddish.[313] These terms have entered the slang of Bulgarian and Greek in the forms *gádže* and γκόμεν-α(F)/-ος(M) / *gómen-a*(F)/*-os*(M), respectively, meaning '[extra-marital] intimate person' (Igla 2018).[314]

### 4.3.10 Isosemy

As noted in §4.3, in contact situations, the semantic structure of words and phrases can come to converge, with the semantic range of a word in one language copied onto a corresponding word in another language, thus extending the range of the word in the copying language. We adopt the term "isosemy" for such equivalence relations on the semantic side holding among items in languages in contact (see footnote 83). For example, Alb *burim* has both the concrete sense of 'spring (of water)' but also the more metaphorical sense of 'source of information,' a range of usage which exactly matches Grk πηγή and BSl *izvor*. In the Greek of Southern Albania, the verb αγαπώ, which means 'love' throughout the Greek-speaking world, has come also to mean 'want,' thus matching the range of Albanian *dua*, which has the same two meanings; thus the question τι αγαπάτε when asked by a server in a café means 'What do you want?' not 'What do you love?'). Macedonian here uses *saka* in both meanings, so that *što sakaš* can mean 'What do you want?' but also 'What do you love?'; interestingly, Bulgarian distinguishes the two: *kakvo iskaš* means 'What do you want?' whereas *kakvo običaš* means

313 This usage is absent in Judezmo (Varol Bornes 2008: 340).

314 See also Kacori et al. 1984: 32 on ethnophaulisms in the secret languages of southwestern Bulgaria.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'What do you love?'. The polysemy of *saka* in Macedonian and thus the convergence with Albanian is therefore an innovation. Similarly, to take another example from the Greek of southern Albania, μηχανή means 'car,' as opposed to 'machine, apparatus; motorcycle' in the rest of Greek, a meaning shift which can be attributed to influence from the somewhat similar-sounding Albanian *makinë* 'car.'[315] Additionally, in the Greek of Palasë in southern Albania, ψημένος, the mediopassive participle of ψήνω 'bake, roast' and thus literally 'cooked, roasted,' can mean 'mature,' just like the Albanian participle *pjekur* (cf. *pjek* 'I bake,' see Joseph et al. 2019), and note also the Macedonian participle *pečen* 'experienced' (cf. *peč-* 'cook, roast').

A somewhat extensive case of isosemy is seen in the various uses of the verb 'open.' In Albanian, Macedonian, and Turkish 'open' (Alb *hap*, Mac *otvori,* Trk *aç-*) is used in reference to turning on lights, lighting fires, lighting ovens, and such, and the meaning with lights is found as well with Greek ανοίγω. Moreover, there is convergence in a metaphorical sense of this verb: Sandfeld 1930: 7, for instance, cites the parallel phrase for what is literally 'my appetite has opened,' meaning 'I have a healthy appetite,' Blg *otvori mi se ištah*, Rmn *mi s'a deschis pofta*, Alb *m' u hap ishtai,*[316] Grk άνοιξε η όρεξή μου, Trk *iştahım açıldı*. Moreover, metaphorical uses extend to derivatives; in particular, the participial forms, Grk ανοιχτός, Mac *otvoren*, and Alb *hapur*, as well as Trk *açık* are used for indicating shading of colors, so that 'light blue' is etymologically "open(ed) blue." Importantly, this usage is absent from earlier stages of Greek or Slavic, so it is reasonable to assume that contact with Turkish is the basis for this use. Albanian and Macedonian seem here to show the most parallelism in the uses of 'open.'[317]

It can be hard to determine both the directionality of influence and the paths of diffusion in some instances, although the occurrence of the Turkish Arabism *iştah* is a smoking gun when it comes to 'appetite'; still, the fact of convergence is undeniable and it is difficult to dismiss contact as the reason for the convergence. Moreover, these examples can be multiplied across all of the Balkans,[318] and similar convergences involving phraseology can be seen in §4.3.10.1 below (cf. also Papahagi 1908).

These instances of isosemy involve content words, but there are cases that involve function words and thus border on the grammatical. One that has been

315 Alb *makinë* is a borrowing from Italian *macchina* 'car' (also 'machine'). From an etymological standpoint, of course, these words go back to Ancient (Doric) Greek μᾱχανᾱ́ 'device, machine' (via Latin *machina*), which in its Attic-Ionic form, μηχανή, is the source of Modern Greek μηχανή. What is relevant here is not the etymology but the similarity in form and meaning between the Albanian word and the Greek word.

316 Albanian *ishta* 'appetite' is an old borrowing from Balkan Slavic that is now dialectal or obsolete; Sandfeld says that *oreksi* (M.DEF; INDF *oreks*), a borrowing from Greek, is also possible here; this Greek loan is the usual form now.

317 Eric Hamp, in talking about this verb with us (p.c., December 6, 2002), said that *hap* has "all of the 'Balkan' senses of 'open.'"

318 See Sandfeld 1930: 36ff. for more examples of this sort, as well as Feuillet 2012, Kyriazis 2012a, and Weinreich 1968:48ff., in some instances with the newer contact-induced meanings ousting the older sense entirely.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.20 *WHERE/WHITHER in the Balkans*

| | |
|---|---|
| Grk | πού |
| Alb | ku |
| Blg | kăde/gde |
| Mac | kade; kaj (COLL) |
| Rmn | unde |
| Aro | iu |
| Megl | iu |
| Rmi | kaj |

mentioned variously in the Balkanistic literature at least since Seliščev 1925 is the convergence of locational and directional meanings for the question word 'where?', thus both 'in which place?' and 'to which place?' (i.e., WHERE versus WHITHER) (see Table 4.20).

Sandfeld 1930: 191–192 mentions this convergence, but he notes that it occurs elsewhere in the Romance languages and can be seen as well in Vulgar Latin; Modern English can be added to the list of languages with such a multiple use of 'where.' The Greek usage could be due to Latin influence, but while Sandfeld is willing to say that the usage in Albanian and Balkan Slavic is due to contact, he is uncertain as to whether it is influence from Romance or from Greek.

As this last example shows, it is clear that some of these developments can be found outside of the Balkans. The 'want'/'love' nexus, for instance, is found in Spanish, where *quiero*, originally 'want,' also means 'love.' Such is the case also with the meaning shift seen with Greek θέλω 'want,' which at some point in the Postclassical period took on the meaning 'need,' as in το στιφάδο θέλει αλάτι 'the stew needs (lit., 'wants') salt.' The same development is found in the use of Macedonian *saka* and Aromanian *va* as in Mac *saka mnogu odenje do Bitola* = Aro *Va multu imnari pănã Bituli* (Markovikj 2007: 166) 'It is necessary to make a lot of trips to Bitola (lit., 'it.wants much going to Bitola'). This shift has an intriguing parallel in Albanian, where *duhet* '(it) needs' is formally the mediopassive form of *dua* 'want,' and in Bulgarian where *iska se* 'be required' is the intransitivized form of *iskam* 'want.' Such usages of 'want' as 'need' occur elsewhere in Balkan Slavic and Balkan Romance. A similar meaning is seen in the English adjective *wanting*, as in *His response was found wanting* (i.e., 'needing something more'). In this last case, one needs also to take note of Italian *volere* 'want' and its derivative *volerci* 'need, require' (with locational element *-ci*), suggesting that Italian or perhaps even Late Latin influence might have been involved in some of the languages.

These extra-Balkan parallels raise the specter of the shifts in meaning simply being natural changes that any language can undergo without any contact influence.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

While we acknowledge this possibility, we do not see it as a compelling counter-argument; even if such were the case, the convergence is real and contributes to the sense of sameness that one sees, and which speakers seem to feel, among the languages of the Balkans in on-going contact situations. Moreover, one could argue that it is contact that allows the natural shift to take hold in any particular language.

#### 4.3.10.1 Phraseological Isosemy

The focus of discussion up to this point has largely been on words, though there have been a few parallels brought to light involving more than one word, for instance the Verb-*NOT*-Verb phrasal parallel (§4.1), the consideration of whole-word reduplication (§4.3.7), and even the various uses of 'open' where the combination with particular objects is at issue (§4.3.10). Such cases make it clear that the parallels in the Balkan lexicon are not restricted to single words (as already observed by Papahagi 1908 and Gilliat-Smith 1915/1916). As might be expected, given the extent to which various features of the Balkan languages match up, there are numerous parallels that extend beyond the level of the individual word to phrases and sentences. In fact, Sandfeld 1930: 205 says that the phraseological parallels "are so numerous that one would scarcely exaggerate in saying that it is rather the exception when these languages differ completely from the phraseological point of view."[319]

These parallels are striking and involve more than just borrowed material; rather, they are essentially calques, showing the same conceptual structure but built with lexical material from each language on its own. And, as noted first in §3.2.1.7 and §3.4.2.2, and reiterated above in §4.3, such phraseological calques have a special value here in that they provide prima facie evidence for bilingualism – there could not be the word-for-word/morpheme-for-morpheme glossing in a calque without bilingualism; since there was no overt classroom-style learning of the other language, the existence of extensive calquing must reflect a situation where natural acquisition was going on, due to interactions on a day-to-day basis (or intermarriage). Moreover, the interactions that these calques document reflect shared experiences that the speakers of the various languages could draw on, so that there is a cultural component to them as well, beyond the purely linguistic.

In some instances the source is known, but in others, as noted elsewhere (§4.3.10), the source and/or the directionality of the diffusion cannot be determined. In such cases, the fact of a convergence alone is sufficient to make the parallels interesting, as they are necessarily based on a matching of surface material

319 "*Elles sont si nombreuses qu'on n'exagéra guère en disant que c'est plutôt l'exception quand ces langues se diffèrent complètement au point du vue phraséologique*." Seventy-five years later, in her comparison of Bulgarian and Romanian phraseology, Kaldieva-Zaharieva 2005: 360 quoted Sandfeld's observation as borne out by her work, but added that some parallels were shared with other Romance and Slavic languages, attributing this to the influence of Christianity and Bible translations. These two observations can be taken as framing two key issues in Balkan phraseology: on the one hand, its specificity, on the other hand, its cultural links beyond the Balkans as well as typological universals.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

in one language with corresponding pieces from another language. Determining the source of "whither" and "whence" convergence is less important than the recognition that linguistic material passed between the speakers involved (cf. Ilievski 1973).

The material that is available on this topic is considerable and rich, with important studies by Papahagi 1908 (supplemented by Çabej 1936), Jašar-Nasteva 1962/63, Ikonomov 1968, and Djamo-Diaconiţa 1968, as well as material to be found in Sandfeld 1930 and the observations in Gilliat-Smith 1915/1916; see also Markov 1977 and Thomai et al. 1999.

In what follows, we can give only a sampling of the types of parallels to be found, and the sorts of categories, i.e., shared experiences, they show. Given the age of some of these sources, and the fact that they documented usage from over a century ago when a greater percentage of the population of the Balkans than now lived in villages, some of these expressions are no longer current or may even be archaic or quaint, at least in some locales. Their value to Balkan linguistics, however, is not diminished by these subsequent developments.

We organize the material into two main groupings, covering on the one hand shared idioms, and on the other shared expressions from various aspects of daily life.

#### 4.3.10.1.1 Idiomatic Expressions

Thomaj et al. 1999 supply c. 5,000 more or less shared idioms in Albanian, Bulgarian, Greek, Romanian, and the former Serbo-Croatian. Absent from their compilation is a very striking set of expressions in the Balkans that involves verbs of ingesting, most notably the verb for 'eat,' which, when occurring together with various objects, forms combinations that refer to some negative event or consequence. These are all either constructs loosely based on, or direct calques of, Turkish models.[320] For instance, 'eat' + 'a blow' means 'get a beating,' where the model is Turkish *kötek* 'blow' + *yemek* 'to eat.' Here are some examples from various languages (4.16):

(4.16) Trk: yağmur yemek 'get soaked' (lit., 'rain eat')
bok yemek 'say something stupid' (lit., 'shit eat')
kötek yemek 'get a beating' (lit., 'a.blow eat')

320 Although Turkish is clearly the immediate source of these idioms based on 'eat' in the Balkans, they most likely did not originate in Turkish and in fact Turkish may be the western edge of a chain of contact by which 'eat'-based negative idioms spread across Eurasia from more easterly sources. This scenario is suggested by the occurrence of such idioms in Korean (e.g., *geob* 'fear' + *meogda* 'eat' => 'be afraid,' *miyeoggug* 'seaweed soup' + *meogda* => 'flunk an exam', *kongbab* 'bean-and-rice' + *meogda* => 'do time in jail'), in languages of Oceania (e.g., the Bislama dialect of Melanesian Creole has 'eat' in idioms in a wide range of meanings involving suffering, including "EAT" + "HAND" for 'to get punched'), and, as the likely proximate source for Turkish, in Persian (where, according to Family 2009, the verb *xordæn* 'eat' occurs with a variety of objects to refer to negative events, e.g., *pa xordæn* 'foot' + 'eat' => 'be stepped on' or *šæmšir xordæn* 'sword' + 'eat' => 'be stabbed by a sword'). Somewhat closer to home, as far as the Balkans are concerned, the Greek "eat wood" idiom has spread to the Italian of Corfu (*mangiare legnate*), according to Cortelazzo 1948: 33.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Mac: jade kjotek ‘get smacked’ (lit., ‘eat smack/blow’)
jade stap ‘get a beating’ (lit., ‘eat stick’)
jade dožd ‘get soaked’ (lit., ‘eat rain’)
jade gomno ‘say something stupid’ (lit., ‘eat excrement’)
Grk: τρώγω ξύλο ‘get a beating’ (lit., ‘eat wood’)
φάγαμε γκολ ‘we had a goal scored against us’ (lit., ‘eat.PST.1PL goal’)
Alb: ha baltë ‘suffer badly’ (lit., ‘eat mud’)
ha dru ‘get a heavy beating’ (lit., ‘eat wood’)
ha dajak ‘get a heavy beating’ (lit., ‘eat a club/cudgel’)
ha mut (lit., ‘eat shit’) ‘talk nonsense’ (*mos ha mut* ‘shut up’, lit., ‘don’t eat shit’; cf. an Albanian internet meme “KEEP CALM AND MOS HA MUT”; for the original “KEEP CALM AND CARRY ON”).

A shift from a basic meaning of ‘eat’ to ‘take in’ (“ingest” in the broadest sense) seems to be involved here, and while it is a natural enough shift to allow for independent origin in each language, the convergence of particular types of direct objects accompanying this verb and the negative meaning associated with the combination, so that EAT has become SUFFER, marks these as likely calques on the model provided by the Turkish construction. Some of the languages show extensions of this usage that are not strictly speaking calques or based on Turkish; Greek, for instance, has τις φάγαμε από τους Ιταλούς ‘we lost to (suffered a loss by) the Italians’ (lit., ‘them.WEAK.ACC.F.PL eat.PST.1PL from/by the.M.PL.ACC Italians.ACC’), with an unspecified (but feminine plural) weak object pronoun, thus “we ate those-things (e.g., losses) at-the-hands-of the Italians.”

Another characteristic idiomatic use of a verb of ingesting is seen with the verb for DRINK, which is (or used to be) used as well with ‘cigarette’ or ‘tobacco’ (etc.) as the object, giving the meaning ‘smoke,’ e.g., Alb *pi cigarë*, Grk πίνω τσιγάρο, Aro *beau tsigară*, Mac *pie cigari* [PL], Rmi *piav tsigaro*/*tutuno* (‘tobacco’), all ultimately based on Trk *sigara/tütün içmek,* although the usage, as just indicated, is now considered old-fashioned or obsolete in some of the languages.[321] Despite its obsolescence for ‘smoke’ in some Balkan languages, DRINK still figures in a usage with ‘pill’ as the object: Grk πίνω χάπι, BSl *pie [h]apče*, Alb *pi hap*, Rmn *bea hap*.[322] In Romani, *ha* ‘eat’ also serves as the basis for ‘understand,’ normally as a derived form, e.g., *haljovel*.

321 Like the EAT-based idioms, this one may have a more easterly origin. Sanskrit has a compound *dhūma-pa-*, in which the second part is based on the verb *pā-* ‘drink,’ meaning ‘drinking in or inhaling of smoke (*dhūma-*).’ This usage is also found in seventeenth-/eighteenth-century Dutch, and van der Sijs 2001 notes that other verbs were also used with tobacco in early modern Dutch, including ‘blow,’ ‘suck,’ ‘guzzle,’ and Modern Dutch has many idioms employing ‘drink.’

322 It is worth noting that Sandfeld 1930: 36ff. gives many such expressions, some of which are still current and others of which are not. Some of the changes that have occurred in the century or so since the appearance of the publications on which Sandfeld 1930 is based can be attributed to ordinary language change, while others are connected to the various effects of the rise of standard languages. The same can be said for Papahagi 1908.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Yet another Turkish-based expression is the use of the verb 'know' with a language as a complement, either as an object or in an adverbial form, as the unmarked colloquial way of saying 'to (be able to) speak a language' (4.17):[323]

(4.17) Alb: e di shqip? 'Do you know (= speak) Albanian?'
(lit., 'it know.2SG Albanian')
Grk: ξέρεις ελληνικά? 'Do you know (= speak) Greek?'
Rmi: džane[s] romane[s] 'Do you know (= speak) Romani?'
Mac: znaeš po-našinski 'Do you know (= speak) our (language)?'
Trk: türkçe biliyor musun 'Do you know (= speak) Turkish?'
(= "the.Turkish.way(ADV) know (*bil-*) Q-you")

Although these examples are verb-centered, calques need not be so. Sandfeld 1930: 120 notes the extraneous use of 'all' accompanying the preposition 'with' in Macedonian, Balkan Romance, and Albanian, a pattern he attributes to calquing on Albanian as the model, (4.18):

(4.18) Alb me gjithë priftiu 'with the priest' (lit., 'with all the.priest')
Mac sose baltija 'with the axe' (lit., 'with.all the-axe')
Aro cu tut căpitanlu 'with the captain' (lit., 'with all the.captain')
Rmn cu cal cu tot 'with the horse' (lit., 'with horse with all')

These widely distributed expressions have readily determinable sources. Less certain, but no less interesting and important are expressions that are broadly represented but without an obvious path of diffusion, as well as other more localized idiomatic convergences. A few choice examples include the following, from Sandfeld 1930 and Papahagi 1908 (unfortunately neither of these sources included Romani or Judezmo), (4.19–4.28):

(4.19) 'hold your tongue' (lit., 'gather your tongue/mouth')[324]
Alb mbledh gojën ('mouth')
Aro adună-ţi gura ('mouth')
Grk μάζεψε τη γλώσσα σου ('tongue') (Sandfeld, 112)

(4.20) 'run for your life!' (lit., 'flee that we flee!')
Alb ikëni të ikëmi
Aro fudziţĭ s-fudzim
Blg bĕgajte da bĕgame (pre-1944 orthography)
Grk φεύγετε να φεύγουμε (Papahagi, 129)

323 The Turkish model has a derived adverbial with -CE for the language known/spoken; Macedonian can also use 'speak,' i.e., Mac *zboruvaš* (*mutatis mutandis*, e.g., *zborviš*), as can some of the other languages. Albanian *shqip* is etymologically an adverb (see §1.2.3.1) but here appears to be a noun since there is a weak object pronoun (*e*) ostensibly co-referencing it (though the *e* is possible also with overtly marked adverbial forms as in *e di anglisht* 'Do you know English,' where the noun form for 'the English language' is *anglishte*). However, Alb *fol shqip* means both 'speak Albanian' and 'speak clearly.' The Greek ελληνικά is ambiguous in form between an adverb (as if "Greek-ly") and a neuter plural nominalized adjective (i.e., "the Greek things").

324 Albanian *gojë* can also mean 'speech,' and Greek γλώσσα also means 'language.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(4.21) ‘once and for all’ (lit., ‘one and good’ (F))
Alb një edhe mirë
Aroună ş-bună
Rmn una şi bună
Grk μια και καλή (Papahagi, 134)

(4.22) ‘a complete ass’ (lit., ‘a donkey and a half’)
Alb gomar e gjysmë
Blg magare i polovina
Rmn un măgar şi jumătate
Grk ένας γάιδαρος και μισός (Papahagi, 135)

(4.23) ‘we are very good friends; we’ve been through thick and thin together’ (lit., ‘we have eaten bread and salt together’)[325]
Alb bukë e kripë hëngrëm bashkë
Aro sare ş-pîne mîcăm deadun
Rmn a mînca pîne şi sare împreună
Grk ψωμί και αλάτι φάγαμε μαζί (Papahagi, 151)

(4.24) ‘I prophesize’ (lit., ‘I throw at the stars’)
Alb e heth ndë ūj [*sic*]
Aro aruc tu steale
Rmn arunc în stele
Grk ρίχνω στα άστρα (Papahagi, 157)

(4.25) ‘daily’ (lit., ‘day with [the] day’)
Alb ditë me ditë
Aro dzuă cu dzuă
Rmn zi cu zi
Grk μέρα με τη μέρα (Papahagi, 159)

(4.26) ‘undoubtedly; pointlessly’ (lit., ‘without word”)
Alb pa fjalë
Aro fără zbor
Rmn fără vorbă
Grk χωρίς λόγο (Papahagi, 167)
Mac bez zbor

(4.27) ‘without a doubt’ (lit., ‘without other’)[326]
Alb pa tjetër
BSl bez drugo
Rmn fără de alta
Grk χωρίς άλλο (Sandfeld, 210)

325 Compare the stricture about treating a guest right from the Albanian code of behavior, the *Kanun* of Lek Dukagjin, Chapter 96, §608, that reads *Mikut do t’i bâhet nderë: ‘Bukë e krypë e zêmer’* (‘For the guest, honor must be made: “Bread and salt and the heart”’). Bread and salt are likewise important symbols of hospitality in Slavic-speaking cultures. In Bulgarian, however, *hljab i sol* is also the idiomatic equivalent of English *bread and water*.

326 For this example, as Sandfeld himself recognizes, there is also an Italian parallel, *senza altro* and a French one, *sans autre*, both with the same form and the same meaning (literal and otherwise); he does not feel that this vitiates the case for calquing in the Balkans, and suggests such an instance “*n’ont qu’une valeur secondaire comme preuve de l’unité linguistique des Balkans*” (‘has only a secondary value as a proof of the linguistic unity of the Balkans’). Presumably, the ultimate source lies in Romance or even Latin.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(4.28) 'shoot a rifle' (lit., 'throw a rifle')

| | | |
|---|---|---|
| Alb | shtij dufeki | |
| Aro | arunc tufek'a | |
| BSl | hvărljam/frlam puška (Blg/Mac) | |
| Grk | ρίχνω τουφέκι | |
| Trk | tüfenk atmak | (Sandfeld, 93) |

Among the calques that are limited in representation across the languages are the following, showing different pairings of languages involved in the convergence (4.29):[327]

(4.29)

| | | |
|---|---|---|
| a. Grk | το ξέρω απ' έξω | 'I know it by heart' (lit., 'it.N.ACC.SG know.PRS.1SG from outside') |
| Rmi | džanav les avral | 'I know it by heart' (lit., 'I.know it from outside') (Agía Varvára, cf. Messing 1988: 61) |
| b. Grk | παίρνω κάποιον τηλέφωνο | 'call someone on the phone' (lit., 'take.PRS.1SG someone.ACC.SG telephone.ACC.SG') |
| Alb | marr dikë në telefon | 'call someone on the phone' (lit., 'I.take someone.ACC on phone') |
| c. BRo | ună stămînă dao | 'one or two weeks' (= Aro; lit., 'one week two') |
| BSl | eden den dva | 'one or two days' (= Mac; lit., 'one day two') |

Such expressions serve as important reminders that even those with wide representation surely involved language-by-language (really, speaker-to-speaker) diffusion and presumably at some point were restricted to perhaps as few as two languages.

While the calques discussed so far have involved expressions, there are also calques at the level of word-internal structure.[328] Thus, the Turkish compound *alış-veriş* 'commerce,' lit., 'taking-giving,' has been borrowed as such into some of the languages,[329] e.g., Alb *allishverish* 'business deal, commerce; dirty business, fraud,' Grk αλισβερίσι 'commercial dealings.' However it is also calqued into a compound with appropriate recipient language pieces: Alb *dhënë-marrë* (cf. *dhV-*, suppletive root of 'give,' *marr-*, root of 'take'), Grk δοσο-ληψία (cf. δο- 'give,' ληπ- from root λαβ- 'take'), Blg *zemane-davane* (cf. *zem-* 'take,' *dav-* 'give'), and Rmn *dat-şi-luat* (lit., 'giving-and-taking,' with *da-* 'give' and *lua-* 'take'). Mac *na ti daj mi* (lit., 'here you.DAT give.IMPV me.DAT') 'you scratch my back and I'll scratch yours,' is of a similar type.

These calques are all interesting in their own right but their importance for understanding the Balkan contact situation cannot be overestimated. Still, Friedman 1986c sounds an important note of caution, one that certainly holds here, as with any putative contact phenomenon:

327 Romanian adds an important perspective on two of these. Regarding (a), Romanian has *ştiu (ceva) pe dinafară* (I.know (something) on from.outside) so this expression may have a wider distribution. So also with (c) (from Sandfeld 1930: 154–155) as it involves an expression found also in Romanian, e.g., *un ceas două* 'one or two hours' (lit., 'one hour two'), thus with Balkan Slavic in a small grouping of languages even if more than two languages strictly speaking.

328 See §4.3.10 for word-level calquing of semantic structure.

329 With the characteristic lowering of stylistic level and movement into more pejorative meanings often found with Turkisms, as discussed in §4.4.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Jašar-Nasteva 1962/63 in her excellent work on Turkish calques in Macedonian gives 350 examples, but a number of these are also identical with English usage, e.g., the use of 'fall' to mean 'come/occur' as in *Bajram se paǵa v nedela = Bayram pazara düşer* = 'Bayram falls on a Sunday' (p.130), *svekrvin jazik = kaynana dili* = 'mother-in-law's tongue (a type of plant with long spiny leaves)' (p.122). Given that the English is not likely to be a Balkan calque, the Macedonian expressions cannot be definitely attributed to Turkish without some sort of documentary evidence.

#### 4.3.10.1.2 Shared Experiences – Shared Expressions

Besides the idiomatic expressions just described that often reveal the arbitrariness of the connection between form and meaning in language, there are also expressions that while arbitrary in their own way, nonetheless are rooted in practices and actions of speakers and the societies and cultures they live in and so draw some motivation from them. Such shared experiences allow speakers to reflect in their language ways in which they interact with their culture and with their social environment, that is with other people as they go about their daily lives. We thus focus here on shared expressions that are tied in some way to social and cultural experiences that are common across the Balkans, specifically those rooted in folk culture as well as those with a more mundane, but no less significant, basis.

##### *4.3.10.1.2.1 Shared Expressions Rooted in Folk Culture*

Storytelling is an important part of traditional culture, and the opening of tales is often characterized by a traditional expression peculiar to the genre. The English *once upon a time* is a clear example of such an opening formula. There is a traditional opening of folktales common to much of a region from the southeastern quarter of Europe across Asia as far as the northwest of the Indian subcontinent: *There was [and] there wasn't*.[330] The formula is well represented throughout Turkic, in Iranian and in Arabic, but also in Czech and Hungarian.[331] In the Caucasus, the opening is found in Armenian, Georgian, and Daghestanian but not in Nakh or Northwest Caucasian (Abkhaz-Adyghe). Although the opening occurs in Hindi, it is not attested in Sanskrit, so the source seems to be Middle Eastern or Turkic. In the Balkans, the formula is only partially present. Table 4.21 gives a representative selection of typical folktale openings from the Balkans and from areas where *there was-there wasn't* openings are typical.

As can be seen from Table 4.21, Turkish and Aromanian pattern together. Balkan Slavic has the pattern attested, but the formula in parentheses is more common. Romanian has a positive/negative juxtaposition, but its version translates as 'there was and as never [before].' Albanian juxtaposes two finite verbs forms, but the second is not negated. Greek also has a double juxtaposition, but without a verb. The Romani

330 See Friedman 1999b regarding the choice of Tense-Aspect-Mood-Evidentiality categories in such formulae.

331 Apparently it does not occur in Slovak, which goes with Polish, Ukrainian, and Russian in this respect.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.21 *'Once upon a time' in the Balkan languages and some relevant others*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***bir*** | ***varmış*** | | ***bir*** | ***yokmuş*** | | | | [Turkish] |
| one | exist.NCNFV | | one | not.exist.NCNFV | | | | |
| ***bilo*** | ***ne*** | ***bilo*** | ***(si*** | ***imalo*** | | ***edno*** | ***vreme)*** | [Macedonian & Bulgarian] |
| was.N | not | was.N | REFL.DAT | there_was.N | | one | time | |
| ***tsi*** | ***shi*** | ***ira*** | ***ma*** | ***nu*** | ***shi*** | ***ira*** | | [Aromanian] |
| what | and | be.IMPF.3SG | but | NEG | and | be.IMPF.3SG | | |
| ***ină*** | ***u̯ară*** | ***fost- au̯*** | | | | | | [Meglenoromanian] |
| one | time | be.PTCP-have.3SG | | | | | | |
| ***a*** | ***fost*** | ***odată,*** | ***ca*** | ***niciodată*** | | | | [Romanian] |
| have.PRS.3SG | be.PTCP | once | as | never | | | | |
| ***ishte*** | ***se*** | ***na*** | ***ç'ishte*** | | | | | [Albanian] |
| be.IMPF.3SG | that | us.DAT.it.ACC | what-be.IMPF.3SG | | | | | |
| **μια** | **φορά** | **και** | **έναν** | **καιρό** | | | | [Greek] |
| one | time | and | one | time-period | | | | |
| ***sas-pe*** | ***kaj*** | ***nas-pe*** | | | | | | [Romani][332] |
| was-REFL | that | not.was-REFL | | | | | | |
| ***ulo*** | ***kaj*** | ***ulo*** | | | | | | |
| ***sine*** | ***kaj*** | ***sine*** | | | | | | |
| was | that | was | | | | | | |
| ***sas*** | ***haj*** | ***sas*** | | | | | | |
| was | and | was | | | | | | |
| ***una*** | ***bes*** | ***abie*** | | | | | | [Judezmo] |
| one | time | was [lit., 'had'] | | | | | | |
| ***era*** | ***bwen*** | | | | | | | |
| was | well | | | | | | | |

332 For Romani the examples with *sas* are Vlax, while those with *sine* and *ulo* are Balkan. The form *sine* is a third-person imperfect, while *ulo* is both the masculine past participle and third-person masculine past of *ovel* 'become,' which supplies various suppletive forms for 'be.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Table 4.21 (*cont.*)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ***bir*** | ***var*** | ***idi*** | ***bir*** | ***yox*** | ***idi*** | | [Azeri] |
| one | exist | was.CNFV | one | not.exist | was.CNFV | | |
| **bir** | **bar** | **eken** | **bir** | **žok** | **eken** | | [Kazakh & Kirghiz] |
| one | exist | was.NCNFV | one | not.exist | was.NCNFV | | |
| **bir** | **bor** | **ekan** | **bir** | **yoq** | **ekan** | | [Uzbek] |
| one | exist | was.NCNFV | one | not.exist | was.NCNFV | | |
| ***iq'o*** | ***da*** | ***ara*** | ***iq'o*** | | | | [Georgian] |
| be.AOR.3SG | and | not | be.AOR.3SG | | | | |
| ***iwk'un*** | ***ur,*** | ***q:aiwk'un*** | ***ur*** | | | | [Lak] |
| be.PaGe | is | NEG.be.PaGe | is | | | | |
| ***zow-n,*** | ***zow-n-ānu*** | | | | | | **[Tsez]** |
| be-UW | be-UW-NEG | | | | | | |
| ***bak*** | ***bak-e*** | | | | | | **[Udi]** |
| be.PRF.3SG | be.PRF-NEG | | | | | | |
| ***yeki/yake*** | ***bud*** | ***yeki/yake*** | ***nabud*** | | | | [Persian & Tajik] |
| once.it | was.AOR | once.it | not.was | | | | |
| ***kaan*** | ***ya*** | ***ma*** | ***kaan*** | | | | [Arabic] |
| was.3SG | or | NEG | was.3SG | | | | |
| ***vahām*** | ***gayā*** | ***thā*** | ***aura*** | ***vahām*** | ***nahīm*** | ***thā*** | [Hindi] |
| there | gone | was | and | there | not | was | |
| ***linum*** | ***e*** | ***chi*** | ***linum*** | | | | [Armenian] |
| be.PRS.PTCP | is | NEG | be.PRS.PTCP | | | | |
| ***egyszer*** | ***volt,*** | ***hol*** | ***nem*** | ***olt*** | | | [Hungarian] |
| once | be.IMPF.3SG | where | not | be.IMPF.3SG | | | |
| ***bylo*** | ***nebylo*** | | | | | | [Czech] |
| there.was.N | there.wasn't.N | | | | | | |

data are especially varied, as is appropriate for its dialectal distribution. The *was-wasn't* type is actually best attested in Vlax dialects, especially North Vlax. A double *was* with a complementizer (similar to the Albanian) or conjunction as connector is well attested for both Vlax and Balkan dialects, although many (perhaps most) tales in Balkan Romani begin with the third-person imperfect of 'be' followed by *jekh* 'one/indefinite article' plus a substantive (*thagar* 'king,' *Rom* 'Rom,' *Xoraxaj* 'Turk,' *phuro* 'old man,' etc.). This is similar to the Meglenoromanian opening 'one time there was.'[333] As it turns out, therefore, the introductory formulae for folktales in the Balkans illustrate well what Hamp 1989a identified as *differential bindings.* The *was-wasn't* type – clearly of Ottoman origin in the Balkan context – is present, but other developments are some form of reduplication (as in Albanian, Greek, and Romani) or a different sort of positive plus negative (as in Romanian). At the same time, the Czech and Hungarian formulae suggest differential paths of spread and retention for this particular formula.

There are also many shared conceptual structures in proverbial expressions that can be found across Balkan languages, as collected by Djamo-Diaconiţa 1968 and Ikonomov 1968, where one finds parallel wording and phrasing, parallel semantics, and parallel use in parallel situations. Not all are restricted just to the Balkans, but they have value nonetheless in that they help to show the Balkans as a cultural "zone"; Djamo-Diaconiţa writes quite movingly about the "wisdom and the bitter truths that [these proverbs] express as well as their stylistic beauty [which] assure them a large circulation among many peoples in diverse languages" and observes that "also, in the past, over a long period, proverbs were considered guides in daily life, containing legal and moral recommendations." As such, they represent shared cultural experiences and shared semantics, encapsulated in pithy sayings, to which speakers could refer, and respond. There is some looseness in the expressions, in that they are not always point for point identical, but they share all the key elements. In what follows, a few select proverbs from Djamo-Diaconiţa's collection are presented; some may be dialectal or archaic in form, but that is not unexpected, given the material.

For instance, for expressing contempt for the lazy and for those who shirk duties, one can say (more or less) "Not all flies make honey" (4.30):

(4.30) Alb s'bëjnë mjaltë gjithë mizat (= 'not make honey all flies.DEF')
Aro tute muştile nu fac n'are (= 'all flies.DEF not make honey')
Blg Vsjaka muha med ne bere (= 'every fly honey not gathers')
Grk δεν κάνουνε όλες οι μύγες μέλι (= 'not make all the flies honey')
Rmn nu fac toate muştele miere (= 'not make all flies.DEF honey')

and for the need to economize and plan for "rainy days," one can say (more or less) "(Save) white money for a black day" (4.31):[334]

333 The significant part of the Meglenoromanian formula is the use of a marked nonconfirmative (cf. Friedman 1999c) and §6.2.5.4.

334 The use of 'white' is connected to Ottoman and Greek usage. Both Turkish (from Persian) *akça (*StTrk *akçe*) and Greek άσπρα refer to 'white' in the context of coinage, specifically in regard to a silver coin that was a minimal unit of currency (cf. the use of *penny* in various English proverbs and nursery rhymes).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(4.31)
| | |
|---|---|
| Alb | ruaj paran e bardhë për ditë të zezë (= ‘preserve money.DEF.ACC white for day.PL PC black.PL’) |
| Aro | bani albi pentru zile negre (= ‘money white for days black’) |
| Blg | beli pari za černi dni (= ‘white money for black days’) |
| Rmn | stringe bani albi pentru zile negre (= ‘gather money white for days black’) |
| Grk | τ’άσπρα για τες μαύρες μέρες (= ‘the white [coins] for the black days’)[335] |
| Mac | beli pari za crni denovi (= ‘white money for black days’) |
| Trk | ak akça kara gün içindir (= ‘white silver.coin black day for.is’) |
| Rmi | pharne pares miźinav kales dives resav (= ‘white money I.conceal, (for a) black day I.grasp’) [Džambaz]) |

To convey the idea that one has to put in effort to obtain some desired object (or that the one who complains profits from it), a proverb roughly comparable to the English *The squeaky wheel gets the grease* is said, with the following content, more or less, “If a baby does not cry, it will not suck (i.e., nurse)” (4.32):[336]

(4.32)
| | |
|---|---|
| Alb | pa mos qarë një fëmijë, nuk i ep e ëma gjijë<br>(= ‘without not crying a child not to.it gives mother breast’) |
| Aro | ficiorlu cari s-nu plîngă nu suge<br>(= ‘child.DEF if DMS-NEG cries not sucks’) |
| Blg | deteto dogde ne zaplače majka mu ne mu dava da bozae<br>(= ‘child.DEF until NEG cries mother its NEG it.DAT gives DMS it.sucks’) |
| Rmn | copilul pînă nu plînge nu suge<br>(= ‘child.DEF until not cries not sucks’) |
| Grk | αν δεν κλαίει το παιδί βυζί δεν τρώει<br>(= ‘if not cries the child breast not eats’) |
| Mac | duri ne zaplačit deteto majka mu ne mu daat da cicat [= Ohrid]<br>(= ‘even not cries child.DEF mother its not it.DAT gives DMS nurses’) |
| Trk | ağlamıyan çocuğa meme vermezler<br>(= ‘cry.NEG.PROG child.DAT breast give.NEG.PL’) |

And, as a call to vigilance even when there seems to be nothing ominous on the horizon, much like the English *Still waters run deep*, one says, more or less, ‘Water sleeps but an enemy does not’ (4.33):

(4.33)
| | |
|---|---|
| Alb | lumi flen armiku nuk flen/Lumi fle, hasmi s’fle [Ikonomov 1968: 45]<br>(= ‘river.DEF sleeps enemy.DEF not sleeps’) |
| Aro | apa doarmi duşmanul nu doarmi<br>(= ‘water.DEF sleeps enemy.DEF not sleeps’) |
| Blg | voda spi, a neprijateljat ne spi<br>(= ‘water sleeps and/but enemy.DEF not sleeps’) |
| Grk | το νερό κοιμάται ο εχθρός όμως όχι<br>(= ‘the water sleeps the enemy however not’) |

335 Peter Mackridge brought to our attention the following variants of this phrase in Greek (from Politis 1899–1902: Vol.2), άσπρα γρόσια για τις μαύρες μέρες (Ioánnina, lit., ‘white groats [= coins] for black days,’ p. 535) and ασπρόγροσια, μαύρες μέρες (Megiste/Kastellorizo, lit., ‘whitegroats, black days,’ p. 547).

336 Some of these forms are nonstandard or specifically dialectal, e.g., Alb *ep* for *jep,* southern Aro *plîngă*, Mac 3SG *-t* (typical of Ohrid) and the spelling of *davaat* as *daat*. The point here, however, is the parallel expressions.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Mac vodata spijat, a dušmanot nikogaš (né spiet) [= Ohrid]
(= 'water.DEF sleeps and/but enemy.DEF never (not sleeps)')
Trk su uyur duşman uyumaz
(= 'water sleeps enemy sleeps.NEG')

For the bitter truth that love is blind, Bulgarian and Turkish share a proverb (Ikonomov 1968: 139) (4.34):

(4.34) Blg ljubovta e kato muha: i na med kačva, i na govno kačva
Trk sevda sinek gibidir, bala da konar, boka da konar
'Love is like a fly, it lands on honey, and it lands on shit.'

Djamo-Diaconiţa's and Ikonomov's rich collections have many more such examples, some of which are found in just a subset of the languages, and some of which are found outside of the Balkans. Djamo-Diaconiţa's interest is not just in documenting but also determining, to the extent possible, what the source language/culture is. Even in the absence of a clear origin, the point of the convergences in this domain is clear: proverbial expressions provide a particularly transparent case of shared cultural and linguistic calquing leading to shared phraseology.

#### *4.3.10.1.2.2 Shared Everyday Expressions*

Speakers of different languages who nonetheless know, to some extent, the other languages in the linguistic marketplace have a degree of common linguistic ground with other speakers. The common experience of daily life together with common languages brings the opportunity for convergence in the phrases that are part of the "glue" of daily interactions, and this is seen in the Balkans. In a sense, it is the phrasal equivalent of the shared discourse items discussed in §4.3.4, though here with calquing, essentially involving common conceptual structures, instead of replication of material. We survey here a variety of these expressions, including a number of greetings. The topic of greetings in an intense and sustained language contact situation is treated at the level of lexical borrowing in §4.3.4.2.2 with reference to instances of borrowed terms used in greetings. In this section, shared greeting structures are documented, reflecting calques across the languages.[337]

As §4.3.4.3.2 makes clear, attention-getting words have been borrowed in the Balkans, but interestingly there is one such word that has been both calqued and copied: Turkish uses *buyurun*, the imperative (plural) of the verb *buyurmak* 'to command,' as a way of saying to an interlocutor, especially a potential patron in a store or restaurant or the like, "you have my attention." The equivalent expression in Greek is ορίστε, an imperative (plural) of the verb ορίζω, which in Medieval Greek meant, among other things, 'command,' though the Ancient Greek verb and the Modern Greek verb mean 'determine, fix, assign, master,'[338] and the same is

337 See also §4.3.8 on expressions of closeness and deference, closely allied to the politeness one can see at the level of day-to-day interactions.

338 The modern meanings may reflect influence from Katharevousa; note that the form ρίζω* might be expected if the modern form continued the ancient form directly.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

true of Albanian, where *urdhëroj* 'command' in the imperative *urdhëro* (singular)/*urdhëroni* (plural) is used, as Newmark 1998: s.v. puts it, to signal "respectful attention to another person's needs or requests." Macedonian uses *poveli/povelete* (SG/PL), and Bulgarian uses *zapovjadaj/zapovjadajte* (SG/PL). Most of the Balkan languages also borrow the Turkish, sometimes with slight changes in form, e.g., BSl *bujrum* (and *bujrumte*, with the BSl 2PL marker), Alb *bujrëm*, Aro *buiurun*, Jud *buyrun*.[339]

Also in the realm of patronage and custom, Greek and Albanian show parallel structures for asking 'How much is it/how much does it cost?'. Literally, the expression is 'how.much does.it.make': Greek πόσο κάνει and Albanian *sa bën*.

Moving more in the direction of conversational exchanges, for 'what is your name?', as Papahagi 1908: 151 gives it, one finds 'how do they say you?', with the pieces ordered just so ('how you.ACC say.3PL'), although other parallel collocations such as 'how do you call yourself / how are you called' (in brackets below) also occur:

(4.35)

| | | |
|---|---|---|
| Alb | qysh të thonë? | [si qyhesh] |
| Aro | cumu-ţĭ dzic? | |
| Blg | kak te kazvat? | [kak se kazvaš] |
| Mac | kako te vikaat | [kako se vikaš] |
| Rmn | cum îţĭ zice? | |
| Grk | πώς σε λένε? | [πώς λέγεσαι] |

Here Romani uses *sar si to anav*, lit., 'how is your name.' And, Sandfeld 1930: 208–209 points out that the answer to a superfluous question for which there is an affirmative response, thus something like 'of course' or 'naturally,' is the same across the Balkans, lit., 'how not?':

(4.36)

| | |
|---|---|
| Mac | kako [da] ne? |
| Rmn | cum nu? |
| Grk | πώς όχι? |
| Meg | cum nu? |
| Aro | cum [di] nu? |
| Alb | si jo? |
| Rmi | sar na? |

The Greek is often shortened to simply πως?! (lit., 'how?!'). Sandfeld takes this to be a pattern that goes back to Ancient Greek and so presumes that Greek is the model here.

Perhaps the most striking conversational parallels come in the area of greetings. For 'welcome!', one finds 'well (that) (have.)you.come':[340]

339 See footnotes 166, 181, 201, 204 for more on *–m*/*-n* in such borrowed forms.

340 The form here is 2PL, but 2SG is also used. The Turkish greeting occurred in Balkan Slavic and Albanian (with WRT phonology in Macedonian, e.g., 2SG *oždželdın*) but is now obsolete (cf. also Gilliat-Smith 1915/1916); see also §4.3.4.2.2.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

(4.37)

| | |
|---|---|
| Alb | mirë se erdhët |
| Blg | dobre došli |
| Grk | καλώς ήρθατε |
| Mac | dobrodojdovte |
| Trk | hoş geldiniz |
| Rmi | mišto aljan |
| Rmn | bine aţ vănit |
| Aro | ghini vinitu |
| Jud | [seas el] byenvenido |

Here it is worth noting that Bulgarian represents the original Slavic use of the resultative, while Macedonian uses the aorist. Similarly, Romanian uses a perfect while Aromanian uses an aorist. These differences between the two respective closely related languages reflect areal patterns: Romanian with Bulgarian, on the one hand, and Macedonian and Aromanian with Greek and Albanian, on the other. Moreover, Turkish, like Macedonian, uses the confirmative *DI*-past (see §6.2.5), which, along with Greek and/or Albanian, could have influenced the Macedonian and Aromanian usages.[341]

Importantly too, and characteristically Balkan (unlike 'welcome' routines in other European languages), there is an obligatory response on the part of the arriving person, lit., 'Well (that) I/we.have.found [you]'.[342]

(4.38)

| | |
|---|---|
| Alb | mirë se ju gjeta |
| Blg | dobre nameril |
| Grk | καλώς σας βρήκα |
| Mac | dobro [ve] najdov |
| Trk | hoş buldum |
| Rmi | mišto arakhljum |
| Rmn | bine am găsit |
| Aro | ghini vi aflai |
| Jud | [seas el] byen fayado/tropado |

Moving on from welcoming, we find parallels in asking how someone is, where there is one locution that is particularly characteristic of the Balkans:[343]

(4.39)

| | |
|---|---|
| Alb | Çka po bën? Mirë. |
| Aro | Tsi fats? Gine. |
| Blg | Kakvo praviš? Harno. |
| Grk | Τι κάνεις? Καλά. |
| Mac | Što praviš? Arno. |

341 We can note here that Azeri uses the *mIş*-past (*hoşgelmişsiniz*), but this is an innovation connected with the shift of evidential functions to other verbal affixes (Friedman 2018b: 127).

342 Here the 1SG is used, but 1PL is also possible. In Italian, there is a greeting *Ben trovato*, lit., 'well found,' that is superficially like the Balkan response here. However, its function is different, being used when you see someone for the first time in a while, not necessarily as an overt act of welcoming. Moreover, it is not part of a bipartite dialogic routine, as in the Balkans, and in any case, it is not necessarily the visitor or returnee who uses the expression with 'found.' The Judezmo is structured similarly to the Italian, but its usage is Balkan.

343 There are of course other ways of saying this in the various languages, e.g., in Greek, τι γίνεσαι, lit., 'what are.you.becoming?', or Albanian *Ç'kemi*?, lit., 'what do.we.have?', but the interest here is in recurring patterns across the languages. Various forms of 'what's new?' 'how's it going,' etc. also occur. Cf. Hughes 2003 on greeting formalities in Kosovar Geg.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

| | |
|---|---|
| Rmn | Ce mai faci? (Fac) bine. (*mai* 'more; still') |
| Rmi | So kere? Šukar. |
| WRT | N'aparsın? İyi. (= "*ne* 'what' + *yap*- 'do' + *ar* 'present tense' *-sın* '2SG') |

This pattern is distinctly informal and colloquial. Today in the Balkans it is taken as emblematic of Balkan sociality: lit., 'what are.you.doing?' for 'how are you?', for which an adverbial response ('well') is also considered typical.

Expressions of thanks are also subject to numerous contact phenomena as well as changes over time. Thus, for example, the normal Balkan Romani expression is *ov sasto* (lit., 'be healthy' [male addressee]; cf. the use of Aromanian *sănătate* 'health'). This corresponds closely to Macedonian *da si (mi) živ [i zdrav]* 'may you (SG) be alive(M) (me.ETH.DAT) [and healthy],' which is now a bit old fashioned. The Greek να είσαι υγιής 'may you be healthy' is not current in ModGrk, but in a reduced form, a presumed ναείς, the result of a clipping of είσαι 'are' together with the mood marker να, appears to be the source of another Romani expression *nais tuke* 'thanks to.you,'[344] and note also the colloquial Albanian *Rrofsh* 'may you be preserved.' Greek is the direct source of Aromanian *haristo* (Grk ευχαριστώ) as well as Macedonian *spolajti*, mentioned in §4.3.4.3.1, which is from Grk σ'πολλά έτη 'to many years' (older σπολλάτη, which was used as an expression of thanks in Greek). The current expression for 'many years' in ModGrk is χρόνια πολλά, which, however, as noted above, has different uses, mostly congratulatory in nature.[345] Macedonian *spolaj ti* today is considered rural or old-fashioned except in Aegean Macedonian, where it is now emblematic vis-à-vis Standard Macedonian. Moreover, as noted in §4.3.4.3.1, *spolaj* is reinterpreted as a kind of imperative, so that *spolajvi* can be used as a 2PL. Relevant here also is Romanian a *mulțumi* 'to thank,' with 1SG *mulțumesc* '(I) thank you,' a verb derived ultimately from Latin *multus* 'many' via use in the birthday wish and congratulatory expression *la mulț ani*, lit., 'to many years' (Cioranescu 1958–1966: 545).[346] The French *merci,* spelled with medial *-s-*, e.g., Trk *mersi,* is a common colloquial 'thank you' in the eastern Balkans, i.e., Bulgarian, Greek, Romanian, and Turkish. Albanian and BCMS share the etymological notion of 'praise' or 'honor' in their current standard, viz. Alb *falemnderit* (1SG speaker, versus *faleminderit* for 1PL (or "royal *we*") speaker(s)) from *fal* 'offer, grant' + *nder* 'honor,' and BCMS *hvala* from *hval-* 'praise.' The BCMS has been borrowed into Macedonian as *fala*. The semantics are similar to Bulgarian *blagodarja* from the adverb *blago* 'well, kindly, etc.' + *darja* 'I grant, give,' which is also the source of Macedonian *blagodaram.* Modern Greek ευχαριστώ is from an Ancient Greek deadjectival verb based on εὐχάριστος 'thankful,' a participial adjective from εὐχαρίζω 'render thanks' (εὐ- 'well' + χαρίζω

344 Note current Greek νάσαι καλά, lit., 'may you-be well,' where νάσαι is a contraction of DMS να with είσαι, the 2SG present of 'be,' used to express thanks and thus essentially now a synonym for 'thank you.'

345 The form έτη is actually Katharevousa, which indicates that the expression was probably a nineteenth-century one, like the spread of *mersi* from French (see 4.2.1.7).

346 Slavic languages still use the Church Slavonic *mnogaja lěta* 'many years,' and Albanian has *për shumë vjet [gëzuar]* 'for many [happy] years,' but this is generally a birthday or other congratulatory wish, as is also the case with Grk χρόνια πολλά, rather than 'thank you.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'do a favor; please'); like Albanian, a 1PL form, Grk ευχαριστούμε, BSl *blagodarime*, is possible. It is probably the case that the Bulgarian/Macedonian formula is a Church Slavonicism that has its origins in a calque on the Greek.

Expressions of leave-taking are often ritualized with different versions for the person departing and the person staying, but here we simply note that wishing someone a good journey – for which English often uses the French *bon voyage* (although nowadays *safe travels, travel safely, be safe,* etc. have become increasingly common) – in the Balkans generally makes reference to a good road (BSl, BRo, Rmi), or good roads (Alb, Grk, Trk):[347]

(4.40)

| | |
|---|---|
| Alb | rruga e mbarë |
| Aro | calea mbar/cale-ambar |
| Blg | dobar păt |
| Grk | καλό δρόμο |
| Mac | dobar pat |
| Rmn | drum bun (Old Rmn *bună cale[a]*, Papahagi 1974: s.v.) |
| Rmi | šukar drom |
| Trk | iyi yolculuklar[348] |
| Jud | kaminos bwenos |

We can also note here a leave-taking parallel between Albanian *mirë u pafshim*, Aromanian *s'nã videm cu ghine*, and Macedonian *da se vidime za arno* all meaning 'may we see one another well/with good/for [a] good [thing/cause].' The Albanian is a standard leave-taking, whereas the Macedonian is used for a somewhat longer absence.

There is one final phrase worth mentioning here, namely that used in Greek in the past for the game of "peek-a-boo" between older people and babies and children.[349] It is μπούλι μπούλι μπούλι μπούλι τζα!, with the μπούλι part used while the face is hidden and the τζα at the end when the face is revealed. This phrase is now obsolete but is still remembered by some contemporary speakers consulted. The pieces of the phrase have no meaning in Greek, except that τζα, as discussed in §4.3.4.3.2, is an attention-getting element that can be used for an unexpected appearance by someone, cf. Mac and Aro *dza!* in peek-a-boo. Regarding origin, τζα in Greek seems to be a borrowing from Albanian (see §4.3.4.3.2), so it is natural to look to Albanian for the rest of this phrase; as argued in Joseph 2010a, the Albanian verb *mbyll* 'close, close together' provides a suitable source: μπούλι would be the rendering of the 3SG past form (contemporary standard Albanian *mbylli*), so that the phrase would be 'It-closed, it-closed, it-closed, it-closed . . . Here-it-is!'.[350]

347 Cf. English *happy trails*.

348 Turkish *yolculuk* is best translated 'journey,' a derivative of *yolcu* 'traveler,' from *yol* 'road.'

349 We word things this way as the most current term used in this game is κουκού, apparently from the noise a cuckoo clock stereotypically makes when it opens up.

350 The phonology of a connection of μπούλι to *mbylli* is not perfect, especially since Arvanitika would have [i] for the [y] found elsewhere with this verb in Tosk, but it need not be the case that Greek speakers would necessarily replicate the word completely accurately. Moreover, the adjacent dark lateral could have given the vowel some acoustic gravity that would make for an acoustic impression of a rounded vowel.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

The key aspect of this account is that it depends on playful and friendly interactions between Albanian-speaking adults (or older children) and Greek babies in order for Greek speakers to pick up such a phrase. The etymological assessment of this phrase therefore gives some substance to the claims made here about the nature of the sociolinguistic setting for Greek and Albanian interactions, and by extension for other Balkan village interactions in the Ottoman period.

There are of course many other phrasal parallels, beyond those mentioned here, including a large number that originated in Turkish and spread from there into other languages in the Balkans. However, these examples suffice to show the pattern of convergence on the form and internal structure of these isosemous phrases.[351]

### 4.3.10.2 Prepositional Calques[352]

A rather extensive domain for isosemy is found in the various uses that prepositions have across the Balkans. Sandfeld 1930: 191, has observed, for instance, that "*on sait que roum.* de *et alb.* për *sont synonymes dans beaucoup de cas*" ('it is known that Rmn *de* and Alb *për* are synonymous in many cases'). Accordingly, we survey here some of the more salient convergences involving prepositional semantics and usage, with some necessary attention to differences as well.[353]Bortone 2010: 241–246 mentions several such cases. For example, in Albanian, Bulgarian, and Modern Greek, the preposition with the *Grundbedeutung* 'from,' respectively *nga*, *ot*, and από, is used with verbs of knowing in the sense of 'know about,' as in (4.41), with 'She understands/knows about cars' given in these three languages:

(4.41)

| | | | | | | |
|---|---|---|---|---|---|---|
| a. ajo | merr vesh | nga | makinat | | | (Alb) |
| she.nom | take.3SG ear | from | car.PL | | | |
| b. tja | razbira | | ot | koli | | (Blg) |
| she.nom | understand.3SG | | from | car.PL | | |
| c. αυτή | ξέρει | | από | αυτοκίνητα | | (Grk) |
| she.nom | know[PRS].3SG | | from | car.PL.ACC | | |

In this case, 'from' used in this way appears to be an innovation; in earlier stages of Greek, for instance, this sense of '(know) about' was expressed by a different preposition, περί (with accusative case). The usage is found in other Balkan languages as well.

The *Grundbedeutung* 'from' figures in two other innovative uses. In the idiom 'pass by (a place),' Albanian and Greek use *nga* and από, respectively, matching the Turkish use of the ablative case, the prototypical 'from' case; cf. (4.42abc), with

351 Newton 1962 has a few phrasal parallels, though not many greetings; Tannen & Öztek 1977 offer some parallels but mainly in terms of function, and not form per se. Hamp 1989a provides an example from the northern Balkans, noting that BCMS *hvala l[ij]epo* and Romanian *mulţumesc frumos* match well German *Danke schön* for 'thank (you) very much,' in that all have an initial word for 'thank(s)' and a modifier 'nice; beautiful'; he attributes this to the influence of years under Habsburg rule.

352 See §7.9 for parallels in prepositional syntax across the languages.

353 See also §4.2.2.6.1 regarding preposition-like prefixes used derivationally.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

's/he passed by the house' in these three languages. Bulgarian has uses of *iz,* etymologically 'from,' meaning 'go along, around, etc.' that are also said to be calqued on the Turkish ablative, see (4.42de), meaning 'he passes along the street/ up and down the street':[354]

(4.42)
| | | | | |
|---|---|---|---|---|
| a. | Kaloi | nga | shtëpia | (Alb) |
| | passed.3SG | from | house.DEF | |
| b. | πέρασε | από | το σπίτι | (Grk) |
| | passed.3SG | from | the house | |
| c. | ev-den | geç-ti | | (Trk) |
| | house-ABL | pass-PST.3SG | | |
| d. | minava | iz | ulicata | (Blg) |
| | passes.3SG | from | street.DEF | |
| e. | sokak-tan | geçer | | (Trk) |
| | street-ABL | passes.3SG | | |

And, in a more grammatical use, the form with the meaning 'from,' either via a preposition or the ablative case, is used for 'than' with comparatives, as with 'sweeter than honey' in (4.43), in most of the Balkan languages (with a hyphen added to the Turkish and Romani to signal the relevant suffix), though not in Albanian, as shown in (4.44):

(4.43)
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| a. | po-sladok | ot | med | | | | (Blg) |
| | CMPV-sweet | from | honey | | | | |
| b. | posladok | od | med | | | | (Mac) |
| | CMPV.sweet | from | honey | | | | |
| c. | γλυκύτερο | από | το μέλι | | | | (Grk) |
| | sweet.CMPV | from | the honey | | | | |
| d. | mai | dulce | decât | mierea | | | (Rmn) |
| | more | sweet | from.how.much | honey.the | | | |
| e. | avgin-dar | pogudlo | / | daa | gudlo | | (Rmi)[355] |
| | honey-ABL | CMPV.sweet | / | more | sweet | | |
| f. | bal-dan | (daha) | tatlı | | | | (Trk) |
| | honey-ABL | more | sweet | | | | |

(4.44)
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| më | e | ëmbel | se | / | *nga | mjalti | (Alb) |
| more | PC | sweet | than | / | from | honey.DEF | |

'sweeter than honey'

In this case, to judge from the evidence of Classical Greek and Old Church Slavonic, where the genitive case on its own was used with comparatives, the use of

354 Bortone observes that non-Balkan languages show this usage too – cf. Italian *passò dalla casa* 'pass.3SG.PST from(= by).DEF.ART house' – and sees this as suggesting that the Balkan convergence is a coincidence and not a significant contact effect. Our view here, as noted above in §4.3.10, is that regardless of the cause, the fact of superficial convergence alone contributes, for speakers in the Balkan multilingual milieu, to the sense of structural and semantic sameness among the languages they use.

355 Some Romani dialects mark comparatives with *maj*, borrowed from Romanian *mai* (and note *daa* in (4.43e) from Turkish); see §6.1.5.5 on Balkan Romani marking of comparatives, which in general shows a complex distribution.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

the preposition for 'from' in (4.43) is an innovation, and thus plausibly contact-induced.

The preposition with the *Grundbedeutung* 'with' also shows some isosemy across some of the Balkan languages. The signaling of means of conveyance uses 'with' all over the Balkans, as shown in (4.45) for 'by train,' with a hyphen added to the Turkish to signal the relevant suffix:

(4.45)

| | | |
|---|---|---|
| Grk | με το τρένο | (το = DEF.ART) |
| Alb | me tren (INDF) / trenin (DEF) | |
| Blg | s vlak | |
| Mac | so voz | |
| Rmn | cu trenul (DEF) | |
| Trk | tren-le ('train-with') | |

Caution is necessary here, as 'with' is found all across Europe in this usage (with hyphens added for clarity of analysis):

(4.46)

| | |
|---|---|
| Itl | con il treno 'with the train' |
| Swed | med tåg-et 'with train-the' |
| Grm | mit der Eisenbahn 'with the train' |
| Hung | vonat-tal 'train-INS' |
| Estn | rongi-ga 'train-INS' |

Still, the facts of (4.46) make for a parallelism on the surface among the Balkan languages that in itself can be significant (see §4.3.10). With human means of conveyance, Greek and Albanian show a convergence in the use of 'with'; (4.47) shows 'with' in the expression for 'by foot':

(4.47)

| | | |
|---|---|---|
| Alb | me këmb | 'with foot' |
| Grk | με τα πόδια | 'with the feet' |

It is important not to get too focused here just on similarities and possible contact influences, because there are many differences, even in elements that show some convergence. For instance, even though there are some striking parallels in the use of *me* 'with' in Greek and Albanian, as just noted in (4.45) and (4.47), there are also differences. For instance, *me* can be used in Albanian in the expression of arithmetic addition, e.g., *6 me 7 është 13* '6 plus 7 is 13' (Newmark 1998:s.v.),[356] whereas in Greek, συν, a learnèd borrowing (from *Katharevousa*) which otherwise in that register means 'with,' is used in that function. Moreover, some of the parallels are limited in scope and show some differences across the languages as well. For example, for the unit by which a sale is measured, e.g., *sell oranges by the kilo*, Greek uses με 'with,' thus με το κιλό 'by the kilo,' a usage which matches Turkish *kilo ile*/*kiloyla* (with the postposition *ile* 'with' in either its separate-word form or its fused harmonic form); Albanian here does not use *me*, employing instead a different construction altogether, and Balkan Slavic uses the preposition *po* (otherwise, 'after').

356 The borrowing *plus* (from French or English) is more commonly used now.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Assenova 2019, while emphasizing that minority Balkan languages in enclaves in Bulgaria, especially Greek and Albanian, show some innovative uses of prepositions that are not based in contact with Bulgarian or any other language, nonetheless documents some contact-induced shifts in prepositional meaning and usage leading to isosemy. We quote her here with minor editing indicated:

> Specific uses of the Greek preposition από ['from, by' – VAF/BDJ], which are not attested in its corresponding prepositions in Bulgarian [in modern terms, Balkan Slavic – VAF/BDJ] and Albanian, were adopted by the South Albanian and Western Bulgarian [in modern terms Macedonian – VAF/BDJ] dialects, which were in contact with the Greek language. It will suffice to mention only a few of them:
>
> • The spatial meaning of "catching" is realized in the Albanian dialects of Zagorie and Myzeqe with the preposition *prej* 'from' instead of the preposition *për* 'for':
>
> *E zuri prej qafe* (Zagorie) 'He caught him by the neck.' (Totoni 1962: 206)
> *E kap pi veshi* (Seman, Myzeqe) 'He caught his ear.' (Thomai 1961: 109)
>
> as in Greek:
>
> πιάνω απ' το χέρι 'I catch his hand,' δένω απ' το δένδρο 'I bind with the tree' (Thumb 1910: 96);
>
> • The expression of *content, storage capacity* […] affects the government of the verb "fill (full)," under the influence of Greek, where after γεμίζω/γεμάτος 'fill/full' it is από that is used: γεμίζει από παιδιά 'full [*sic*, 'fills' – VAF/BDJ] of [/with – VAF/BDJ] children,' but με/*me* 'with' is also acceptable: στρώμα γεμάτο με μαλλί 'a mattress stuffed with wool' (Tzartzanos 1946: 89–90). The dialect of Goce Delčev (former Nevrokop, South-Western Bulgaria) […] has taken over the Greek structure: *Pazar'e e păl'an' ot narot* for "*pălen s narod"* 'The market place was full of people.' (Mirčev 1963: 109).

Besides sounding the important caveat that even in contact situations languages can undergo their own internal developments, the facts Asenova offers here show the continued functioning of Balkan sprachbund processes in the post-Ottoman period.

### 4.3.11 Ethnographic Vocabulary

By *ethnographic vocabulary* we understand those items generally associated with folklore or traditional culture. At issue are terminologies for traditional practices associated with life cycle events (birth, marriage, death, etc.), the calendar cycle (spring, fall, mid-winter, etc.), genres and motifs in folk literature (e.g., types of songs), folk beliefs (e.g., types of spirits), and so on. These terms have been the focus more of ethnography than of linguistics. They overlap with, but differ from, Trubetzkoy's *Kulturwörter* insofar as Trubetzkoy's term is generally understood to reflect cultural specificities that are not necessarily connected with universal features of human life. Thus, for example, the Turkism *kurban* 'sacrifice' (and, in

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

nineteenth-century Macedonian, also 'eucharist') is a classic *Kulturwort*, but insofar as it is connected with ritual practices, it is also ethnographic vocabulary. Similarly, kinship terms (§4.3.1), insofar as kinship is a cultural construct (Schneider 1968, 1984), are also ethnographic vocabulary. On the other hand, some ethnographic vocabulary does not fit neatly into other categories, and so a couple of examples are given here.

One example of ethnographic vocabulary connected with life cycle events is the term for a woman who has recently given birth. In Macedonian, the native terms derived from *rodi* 'give birth' – *rodilka* and *porodilka* – are less frequent than the Greek-derived *leunka* or *lehonka* (Lj. Risteski 2019: Map 5). Particularly significant in this regard is that, with the exception of a few isolated survivals of native terms, the pattern of distribution moves from southwest to northeast, just like many morphosyntactic Balkanisms. The term also appears in Aro *lehoánă*, and Alb *lehonë.* The source is Medieval Greek λεχώνα 'woman in childbed' (ultimately a derivative of PIE **legh-* 'lie'), an ν-stem formation replacing AGrk λεχώ(ς) 'woman in childbed' (cf. English *lying-in* 'postpartum confinement'). Blg *lehusa* (whence Trk *loğusa* 'idem,' BER III: s.v.) appears to be a more learnèd borrowing from a related form involving a participle of the Classical Greek verb λέχομαι 'to lie' (or λέχω 'lull to sleep').

Another example of Balkan ethnographic vocabulary is seen in the Albanian genre of epic songs called *këngë kreshnike* 'heroic songs,' where *kreshnik* 'hero' is from BCMS *krajišnik* (Orel 1998: s.v.) 'inhabitant of the *Krajina* (lit., 'border region,' today's northwestern Bosnia, formerly known as Turkish Croatia).' As Kolsti 1990 has discussed in detail, Albanian-Slavic bilingualism was a crucial factor in disseminating oral traditions. Thus, while the Albanian *kreshnik* in and of itself is merely a Slavic loan, it points to a much larger practice of multilingualism that characterizes the Balkan sprachbund when placed in its ethnographic context.[357]

## 4.4 Register and Style

Differences in register in a given language are simultaneously among the most universal and the most language specific. Arguably, all languages have an elevated register, formal and informal registers,[358] and various linguistic means of indexing social relations. The most frequent of such means is lexicon, but grammar can also be deployed. Thus, for example, Javanese is famous for its complex grammatical and lexical means for indexing an intricate set of hierarchical social

357 See also Plotnikova 2004, Sobolev 2005–2013, Domoselickaja 2012, and Domosleckaja & Sobolev 2018 for useful sources of material.

358 Bloomfield 1927: 436, for instance, classically observes that speakers of Menomini (an Algonquian language of Wisconsin) had metalinguistic judgments about different levels of discourse: "The Menomini will say that one person speaks well and another badly, that such-and-such a form of speech is incorrect and sounds bad, and another too much like a shaman's preaching or archaic."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

relations (Errington 1988). The so-called T/V contrast of informal versus formal second person pronominal address is a well-known distinction, which has also affected the Balkan languages (see §6.1.4.3). In multilingual societies, code-switching can serve as another means of signaling register shift.[359] In this section, we examine some shared features of register development and shift in the Balkan lexicons. Owing to the fact that Turkish and Romani occupy, respectively, the historically most and least privileged of the Balkan lexicons in terms of pre-twentieth century social hierarchies, their lexicons each occupy specific positions that are pan-Balkan.[360] Other socially determined registers that involve either Balkan language contact or common experiences for the Balkans are also treated in this section.

### 4.4.1 The Position of Turkish

Turkish enjoyed a special prestige in the Ottoman Empire as the language of power, commerce, and urban status. Under the Ottomans, *şehirli* 'town dweller' was a privileged tax category that required a minimum of forty years residency, and knowledge of Turkish was *de facto* a part of acquiring this desirable status (cf. Ellis 2003: 2). Turkish was thus not only the language of the market place and inter-ethnic communication but also the language of urban sophistication and privilege. Thus, for example, it is significant that while speakers of Albanian, Aromanian, Greek, and Romani all code-switch into their native languages in the Macedonian ethnic jokes collected by Marko Cepenkov in the nineteenth century, Jews, like Turks, speak Turkish (Friedman 1995b). As freedom of movement increased in the late eighteenth and throughout the nineteenth centuries, larger numbers of non-Turkish speakers moved to towns and learned Turkish. The result was a flood of loanwords into the vocabularies of these new urbanites (cf. Koneski 1981: 187–189). In the mid-nineteenth century, for example, the Bulgarian writer Ivan Vazov (in Vazov 1955–1957: XIX, 335) described urban Bulgarian as *poluturski* 'half-Turkish.' At precisely the same time, however, new nation-building movements (invariably termed 'renaissances' or 'rebirths' since the ideology of the day required some sort of pedigree for a 'nation' to claim legitimacy) were attempting to establish new forms of identity, utilizing, among other characteristics or social facts, language. While the history of language in each nation-building movement has its specificities, one of the common features of these movements in the Balkans was the rejection and replacement of Turkish vocabulary in the formal registers of these new, standardizing languages. This in turn led to the stylistic lowering of

359 As is well recognized (e.g., Heath 1989; Friedman 1995b; Paz 2018; see also §§3.2.1.6, 4.2.2.6.2), the distinction between a borrowing and a codeswitch is not clear cut, although in the case of some languages that are in the process of shift, predication can be taken as diagnostic (so Matras 2012 for Domari). In situations of stable multilingualism, however, such diagnostics are flawed (cf. Auer & Muhamedova 2005).

360 To be sure, during the Ottoman period, Greek held a high prestige among the *Rum* millet (Greek Orthodox Christians), but as the language of the state, Turkish generally had prestige as the language of power and also of urban identity.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Turkisms and the creation of a shared, informal register.[361] This process, like the ideology behind it, was repeated for each language from the nineteenth until the late twentieth centuries, so that in Greek, Romanian, the former Serbo-Croatian, Bulgarian, Albanian, and Macedonian, the same Turkish word or derivational affix can have the same stylistic effect. Thus, for example, the Turkish agentive suffix *-CI* (see §4.2.2.4) can be used in all the Balkan national languages (Alb *-xhi,* Blg and Mac *-džija*, Rmn *-gi*, Grk -τζής) to form slang terms such as Mac *politikadžija* or Alb *politikaxhi*, both meaning '[corrupt] politician' that can be contrasted to new words, such as Mac *političar*, Alb *politikan* as informal vs. formal register. In some cases, Turkisms are used for everyday physical objects but not for metaphorical extensions, e.g., Mac *tavan* stands for the ceiling in a building, but not, e.g., a price ceiling, for which the Gallicism *plafon* must be used (Friedman 1986c). If the metaphorical extension is pejorative, however, the opposite can apply, e.g., ModGrk είσαι ντουβάρι, lit., 'you are a wall' (ντουβάρι from Trk *duvar*), means 'you are a blockhead,' whereas είσαι τείχος, lit., 'you are a wall' (from AGrk τεῖχος), has only the literal meaning and cannot be used as an idiomatic insult.

Kazazis 1975 illustrates the result of this process of lowering in his discussion of a Turkish grammar of Ancient Greek. For example, Trk *araba* is the normal word meaning 'carriage, cart, vehicle.' It corresponds to AGrk ἅμαξα and ModGrk αμάξι. Modern Greek also has αραμπάς from Turkish, but this word is used only to mean '(ox)cart,' and is pejorative when used as a synonym for 'vehicle.' As Kazazis 1975: 18 observes, for a Greek, seeing αμάξι translated into Turkish as *araba* "would be enough to produce at least a smile," but when the same Turkish word is used to translate Ancient Greek ἅμαξα, "that smile often turns into outright laughter." Kazazis' point here is that the Modern Greek diglossia that produced the elevated formal register of *Katharevousa* in opposition to the everyday conversational *Dimotiki* – an official distinction that dominated Greek discourse from the nineteenth century until the official rejection of *Katharevousa* in 1976 – renders the disjunction between Ancient Greek and the colloquial register of Modern Greek especially salient. While the contemporary prestige of Ancient Greek vis-à-vis Modern Greek differs from that of Latin vis-à-vis Balkan Romance or Old Church Slavonic vis-à-vis Balkan Slavic (or Sanskrit vis-à-vis Romani), owing to differences in the processes of standardization, nonetheless the disjunction for those languages with standard forms is similar.[362]

361 Interestingly enough, some of this same vocabulary that had entered the Balkan languages via Turkish was also the object of exclusion in Turkish language reform when it was of Arabic or Persian origin, e.g., *münasip* (< Arbc *munāsib*), Mac/Blg *munasip*, Alb *mynasip* 'suitable' has been replaced, by *uygun*, *zgoden*, *umesten*, and *përshtatshëm*, respectively (cf. Kazazis 1972).

362 See Kazazis 1975, Friedman 2005c, and Kyriazis 2012a on the related phenomenon of mock Turkisms in Albanian, Bulgarian, Macedonian, and Greek. These are jokes that combine Turkisms to produce a humorous effect by implying that Turkish uses primitive combinations for modern concepts, e.g., Mac *gjurultu kutija* 'radio' (lit., 'noise box'), Alb *surat tapi* 'passport' (lit., 'face certificate'), Grk χαγιάτ μπαστούνι 'umbrella' (lit., 'life stick'), Blg *barut kjufte* 'hand grenade' (lit., 'gunpowder meatball').

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

This transition did not follow a rectilinear path. On the contrary, there were various forms of language ideological resistance. Thus, for example, although Brailsford 1906: 86, whose philhellenism (and anti-Semitism) are reflected in his account with the claim that Turkish in Macedonia had only limited uses, Herbert 1906: 152–153 gives quite a different account of Bulgaria, one worth citing here.[363]

> In spite of all that has happened to the Ottoman Empire during the last two hundred years, Turkish is still of paramount importance. In Bulgaria the increase in the use of the Turkish tongue in daily intercourse was to me one of the most striking features of my recent revisit to the principality after an absence of twenty years. Under Turkish government, the Bulgarian knowing Turkish would speak the tongue only on compulsion, in a court of law, or when talking to an official or a gendarme, and he would, in front of his compatriots, be ashamed of his knowledge. Now it is a distinction and a sign of superior education, much as it is a distinction to know French in Germany and England. Formerly, when a Bulgarian not knowing Greek had to speak to a Greek not knowing Bulgarian, they had to employ an interpreter, if they had not some little common knowledge of French or German; now they use Turkish. The attendants in the inns of the smaller Bulgarian towns, where French and German are not spoken, know Turkish. A Bulgarian speaking to a Turkish subject of the principality is expected to know Turkish. The language has the glamour and the romance of five centuries of distinguished and often noble history, of which the events of the last thirty years have been unable to rob it. On what other supposition can one explain this striking fact that in the so-called Turkish theaters of the larger Bulgarian towns, Bulgarian plays, performed by non-Turks for Bulgarian audiences, are done in the Turkish language?

At that same time, however, Bulgarian writers were using Turkisms to signal uncouth, uneducated, "un-European" characters, thus both illustrating, and contributing to, the current of stylistic lowering that Turkisms were subjected to in all the Balkan languages (cf. Friedman 2010a: 7–9). As indicated above, the same scenario was repeated as each new Balkan standard language achieved acceptance. Thus, for example, a year after the official recognition of Literary Macedonian, Blaže Koneski in 1945 wrote an article in which, among other things, he severely criticized a Macedonian translation of Molière's *Le Tartuffe* for being full of Turkisms, writing: "*Toa znači . . . da go snižis . . . istančeniot poetski jazik na Moliera . . . do nivoto na našeto balkansko, kasabsko, čaršisko muabetenje*." ('It means lowering the refined poetic language of Molière to the level of our Balkan small-town marketplace chit-chat.'); cf. also Ežov 1952: 211; Gołąb 1960; Markov 1955. Similar attitudes are expressed for the other standard languages, e.g., Close

363 Brailsford 1906 described Greek as "more serviceable as a polite or commercial language" (p. 86) and lamented that the Balkans had not been hellenized (p. 107). As recently as 1973, however, the Turkish consul in Skopje reported that he could speak Turkish with almost anyone over the age of fifty, and in Kosovo in 1976 Turkish was still the language of prestige among urban Albanians (VAF field notes). In addition to letting his philhellenism distort his account of Macedonia, Brailsford 1906: 82, 84 expressed his anti-Semitism in the following terms (see also §3.0, footnote 1): "Spanish, oddly contorted and corrupted, is everywhere the language of Macedonian Jews . . . [They] turn the stately Castilian of the Middle Ages into a patois for nasty pleasures and petty gains."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

1974: 119, 154, 199 on Romanian, Kranji 1965 and Žugra & Kaminskaja 2003 for Albanian. Kazazis 1977: 302–303, in his review of Dizikirikis 1975, sums up the Greek version of this attitude:

> [. . .] depending on their origin, loan-words differ as to the degree to which they defile a language. Thus, the Romans, the Franks ('[medieval] West Europeans'), the Venetians, all left their linguistic (read: lexical) imprint on Greek. Those were, however, civilized nations, so that their loan-words into Greek are not much of a disgrace and do not wound the 'linguistic dignity' of the Greeks as Turkish loan-words do (6ff. and passim). The latter are a shameful reminder of the centuries-long abject subjugation of the Greek nation to a culturally undistinguished people, the Turks.

For Greek, the exchange of populations between Greece and Turkey mandated by the 1923 Treaty of Lausanne resulted in a new influx of Turkisms, since many of these Christian refugees were monolingual Turkish speakers, and, for the most part, those who spoke Greek were town-dwellers who also knew Turkish. The status of the refugees, however, reinforced the stylistic lowering of Turkisms. Some Turkisms also spread beyond the boundaries of the Ottoman Empire when territories that had never been under Ottoman rule were united with territories that had in the new nation-state of Yugoslavia (similarly for parts of post-World-War-One Romania that had only briefly been Ottoman territory).

The position of Turkish loanwords in the various Balkan languages by 1989 was essentially that described by Kazazis 1972, i.e., (1) fully integrated, neutral loans; (2) low register, including informal, ironic, and pejorative; (3) historical/epic/archaic; (4) local color/dialectal/specialized lexicon. We have discussed neutral and low-register loans, as well as some historical loans in §§4.2.1.6 and 4.2.2.4 as well as in this section. The Balkan epic/folk poetic register is by its very nature archaic or archaizing, and it is especially hospitable to Turkisms, since these registers are, in their thematics and vocabulary, the product of the Ottoman period (Lord 1960: 305–308).[364] For Greek, this connection is ideologically undesirable, as illustrated by Notopoulos's 1959: 1 attempt to create a seamless connection between Homeric epic through Byzantium to Modern Greek epic by passing over the crucial Ottoman period in silence:

> From the days of Byzantium until recent times Greece has had to fight for survival . . .. [The songs] have instructed the generations in the modern counterpart of the Homeric *aretê* [sic], *leventyá*, the gallant attitude toward life. . . . The occasions for recitation are the many opportunities offered by the church for religious holidays and festivals, . . . and that indefinable mood for joyous expression in sheer living which the Greeks call by that unique word, *kephi*.

What this account fails to mention is the fact that during the period between Byzantium and "recent times," it was the Ottoman Turks who brought to Greek both

364 Cf., e.g., Jašar-Nasteva 1987, Friedman 2012d.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

*leventyá* (λεβεντιά, from Turkish *levend/levent* 'conscript, irregular soldier' as well as 'handsome, strong youth' and 'free, independent, adventurer, irresponsible')[365] and *kephi* (κέφι, from Turkish *key[i]f* 'pleasure, delight, enjoyment, merriment, tipsy, etc.').[366]

The role of Turkisms in registers that are "dialectal" vis-à-vis the standard sometimes functions as emblematic for speakers in those Balkan languages that were standardized in the twentieth century. Thus, for example, speakers from Bitola consider the Turkism *nejse* (Trk *neyse*) 'nevermind, whatever' (lit., 'what it.may.be') to be particularly characteristic of their dialect, and Kosovar Albanian uses many Turkisms where the colloquial Albanian of Albania has already adopted standard (non-Turkish) forms (Hughes 2003). This emblematicity of Turkish as colloquial has two further developments since 1989, one for national standard languages in post-communist countries, the other for languages that were only admitted to official use after 1991, i.e., Romani and Aromanian in the Republic of North Macedonia.[367]

In the case of nation-state standard languages in post-communist countries, Bulgarian, Macedonian, Romanian, and Albanian all experienced the same penetration of Turkisms into registers from which they had formerly been excluded. This was especially true in popular media, which experienced a kind of colloquialization as democratization in which Turkisms served as emblematic. In the case of the former Serbo-Croatian, the break-up resulted in the Bosniak claim to Turkisms as Bosnian (Friedman 2005c), while Croatian pursued its long-standing puristic tendencies, and Serbian and Montenegrin more or less continued the pre-1991 lexical status quo.

In Romani and Aromanian, the forms for the 1994 Macedonian census – which was concerned with economic variables as well as enumerated individuals – provide excellent examples of how colloquial Turkisms can be used as standard even though they do not have this status in Turkish itself. In the questions pertaining to bathrooms and toilets, all those languages with established, elaborated norms used euphemistic neologisms or recent borrowings as their official terminology (P-2, VI.8 and 9 in Zavod za statistika na Republika Makedonija 1996): Mac *banja, klozet*, Alb *banjo*, *nevojtore*, Trk *banyo*, *banyo-ayakyolu*, Srb *kupatilo, klozet.* Except for the Serbian deverbal noun meaning 'bathing place,' all the words for

365 Cf. also Persian *lewend* 'free, independent, adventurer, soldier, servant, laborer, libertine, ignorant, layabout, strumpet, gallant, etc.' Forms of *levend* are found all over the Balkans and as far afield as Hungarian and Ukrainian.

366 The ultimate source of κέφι/*key[i]f* is Arabic *käyif* 'state, humor, mood, good mood, pleasure, high spirits, narcotic, etc.' Derivatives of *key[i]f* have made it all the way to Russian *(kayf*), where the meaning is 'high,' especially in reference to the effects of cannabis products. This word is widespread; cf. BCMS, *ćef*, Alb *qef*, Rmn *chef*, etc., though there is no such loanword in Slovene. Trk *levend/levent* is from Italian *levantino* 'eastern' (Tietze 2016b: s.v.), cf. Eng *Levantine.* There is, therefore, some irony in Notopoulos's equation of Homeric *aretḗ* with 'Levantine,' a word that comes with significantly different ideological baggage.

367 Romani now has a considerably more widespread recognition, but in the Balkan context North Macedonia is still the only country where it is official in both federal and local administrative documents.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

'bath' are Latinate borrowings. The Macedonian and Serbian words for 'toilet' are from the British [*water*]*closet,* while the Albanian and Turkish are neologisms that can be glossed as 'necessarium' and 'bath-footplace,' respectively. The Romani documents, however, used the Turkisms *hamami* and *kenefi*, respectively. *Hamam* is the standard Turkish word for 'bath' but has come to mean 'Turkish bath' or 'public bath,' while the *kenef* is considered vulgar in Turkish as well as in the other Balkan languages (BSl *kenef,* Alb *qenef*, Rmn *cheneaf*, and, though rare today, regional ModGrk κενέφι; for Turkish, the word entered via Arabic [Tietze 2016b: s.v.] and may well have started out as a euphemism that became polluted by its association with a dirty place, cf. §4.3.9). For Aromanian the forms were *hàmami* and *hale*, respectively. The latter, from Turkish *helâ*, appears in Albanian as *hale*, where it is considered colloquial, and in Macedonian as *ale, vale,* where it is a regionalism no longer understood in many areas.

For Judezmo, it is interesting to note that the position of Turkisms among educated élites during the late nineteenth century was subject to the same kinds of negative evaluations as was the case with co-territorial languages that had become or were aspiring to become vehicles for nation-states (Bunis 2023). On the other hand, colloquial Judezmo, as reflected, for example, in humorous texts (e.g., Bunis 1999, 2023), was not affected by such "modernizing" tendencies. And since Judezmo has never had the status of an official language anywhere, it was likewise not subject to the pressures that suppressed and then elevated Turkisms in former communist countries. It is thus the case that Judezmo occupies a unique position vis-à-vis Turkisms in the Balkans between the standardized nation-state languages, on the one hand, and locally recognized minority languages (Aromanian, Romani) on the other.

In sum, the position of Turkisms in the Balkan nation-state languages today is still a shared feature. Having been pushed down stylistically during the course of the nineteenth and twentieth centuries, today, as in the past, they are part of a register that, depending on the context, can be earthy, familiar, and homey or crude, vulgar, and loutish. They can signal positive, old-fashioned values or backwardness. They can index the old-town urban or the isolated rural. These seeming opposites are in fact the Janus-faces of values that are carried by the term *Balkan* itself.

### 4.4.2 The Position of Romani

In the Balkans, as elsewhere, the marginalization of Romani speakers is reflected in the position of Romani elements in the various languages with which it has been in contact.[368] These elements are informal, colloquial, slang and cryptolectal, and taboo. Thus, for example, in Bulgarian *gádže* 'girlfriend' from Romani *gadží* 'non-

368 Leschber 1995 makes the point that there are two types of Romani lexical material in Romanian, that in the speech of formerly Romani-speaking communities that have undergone language shift and that used by Romani speakers when speaking Romanian for expressive purposes. Both types have the potential to be taken up by the broader Romanian-speaking community, but the registers

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Romani woman' is as ordinary, but also as strictly colloquial, as English *pal* 'friend' from Rmi *phral* (but also *pral, pal* in some dialects) 'brother.' Macedonian *džukela* 'street dog, mutt, nasty person' from Rmi *džukel* 'dog' is slang and stylistically lower than standard Macedonian *kuče* 'dog.'[369] We can also note here Mac *kăne* from Aro *cãne* 'dog,' which in Macedonian is used only metaphorically to refer to an unpleasant person. For Greek we have πουρό, πουρός '(very) old, past one's physical prime < Rmi *phuro* 'old.M' (Tzitzilis 2006, which provides important insights into the routes of Armenian lexicon into both Romani and Greek).

Göktaş's 1986 lexicon of the slang of Turkish shadow-play (Karagöz) performers has a large number of Romani elements, e.g., *çori* 'knife' (Rmi *čhuri*), *gaco* 'woman' (Rmi *gadžo/gadži*) 'non-Romani man/woman,' *habbe* 'food' (Rmi *habe* 'food' (cf. *ha-* 'eat')), *kerizci* 'singer' (Rmi *kerisar-* 'carouse'), *Matiz, Matto*, the performers' slang name for the character otherwise known as *Tuzsuz Deli Bekir* (Rmi *mato* (f: *mati*) 'drunk'), *naş* 'go' (Rmi *naš-* 'run away'), *peniz* 'letting someone else speak' (Rmi *phen-* 'say, tell'), *piyiz* 'alcoholic drink' (Rmi *pi-* 'drink'). An interesting item is *todi* 'Gypsy' and its derivative *todice*, which means '[in] the slang of Karagöz performers,' and which, perhaps, comes from Romani *tho[v]di* 'placed, washed' (perhaps a reference to Romani cleanliness practices). Kyuchukov & Bakker 1999 supply twenty-six Romani lexical items from the gay slang of Istanbul, and note the use of Romani lexicon among Turkish musicians, some of whom are monolingual Turkish speakers of Romani origin. Kostov 1970 also discusses Romani elements in Turkish slang. Leschber 1995 gives sixty main items as entries known in modern Romanian, among them *dic!* 'look!' (Rmi *dikh* 'look!, see!'), e.g., *dic la el!* 'get a load of him,' *gagiu, gagică* 'guy, gal,' *matol* 'dead drunk,' *şuriu* 'knife (especially the kind equivalent to our switchblade in form or function)' (Rmi *čhuri* but *śuri* in most dialects in Romania), *a hali* 'eat' (Rmi 3SG *hal*), *a pili* 'drink' (Rmi 3SG *piel*), *zbanghiu* 'unreliable, nuts' (Rmi *bango* M/*bangi* F 'crooked, lame, etc.'). Graur 1934 and Julliand 1952 also discuss Romani elements in Romanian. In the case of Bulgarian, most Romani vocabulary occurs in specialized jargons (Kostov 1956; see §4.4.3), but in addition to *gádže* cited above, we can note from Armjanov 2001 *mató* 'drunk,' *bangija* 'stupid person; jalopy,' and *dikiz* 'sight, observation' with derivatives, now obsolete. This last form appears to have entered via Turkish, where *-iz* is a common suffix on words derived from Romani (cf. the forms from Göktaş cited above) and the form *dikiz* is also attested (Aktunç 1990: 84; cf. also Leschber 2002). Petropoulos's 1993 dictionary of Greek gay slang contains a number of items of Romani origin, e.g., δικέλω 'see,' ντικ 'sight, eye,' also ντικ! 'there you are!';

appear to be different in at least some cases. She notes (pp. 172–173) that speakers who had shifted from Romani to Romanian and who no longer knew Romani nonetheless treated the Romani words as unmarked, whereas for other Romanian speakers they were all highly marked. The same could be true for other Balkan languages, since most are also spoken by formerly Romani-speaking communities that have shifted as well as by communities that have maintained Romani as their home language. This is a matter that requires further investigation.

369 Note that in Romani itself, the feminine *džukli* 'bitch' can also mean 'whore.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

μπαρό[ς] 'fat,' cf. Aktunç 1990 μπαρό 'customer, rich person' (Rmi *baró* 'large, big, important'); μπαγγόλος 'squint-eyed,' μπαγγόλα 'deaf'; χαλ 'food,' χάλω 'eat,' χάλε κούλα 'eat shit!' (cf. ModGrk να φας σκατά!), χουλά 'shit' (Rmi *khul* 'shit'); λούμπα, λουμπίνα, λουμπουνιά, λούμπω 'bottom [passive male homosexual]' (Rmi *lubní* 'whore,' *lubikanó* 'lustful, debauched'); cf. also Kyuchukov & Bakker 1999 on Romani in Turkish gay slang and §4.3.9 on taboo lexicon of Romani origin. See also Leschber 2009ab and the references therein on Romani lexis in other Balkan languages and in reference to the discussion in §4.4.3 below.

### 4.4.3 Slang, Cryptoglossia, Jargon

So-called secret languages are types of registers insofar as they generally consist of lexical items embedded into the grammar of the language from which they are hiding.[370] As in English and many other languages, various kinds of syllable shift, insertion, and word-play are among the techniques used to disguise speech. These types of secret languages are used by various social groups usually defined by age (e.g., children, teenagers, youth) or social or professional category (students, gays, masons, carpenters, musicians, the Karagöz players mentioned in §4.4.2, etc.). From a Balkan linguistic perspective in this section, the interesting point comes when these languages borrow lexical items from other Balkan languages, as is the case in a variety of professional jargons, which, owing to social factors, are usually limited to men from a specific region or village. Thus, for example, the secret languages of North Macedonia borrow from Albanian, Aromanian, Greek, Turkish, and Romani, although Jašar-Nasteva 1953abc makes the point that Albanian elements are especially prominent.[371] The secret mason's language of Goce Delčev (formerly Nevrokop in Pirin Macedonia, now the Blagoevgrad [Gorna Džumaja] district of SW Bulgaria) has a similar lexical profile (Karastojčeva 2010). Many such expressions are also shared with masons' secret languages in the Rhodopes (Keremedčieva 1995) and central Bulgaria (Ivanov 1974). The material in Kacori et al. 1984 gives similar evidence of the importance of Albanian for secret languages in southwestern Bulgaria in general. In Albania, secret languages in southern Albania tend to borrow from Aromanian, Greek, and Macedonian (Shkurtaj 2004; Sh. Demiraj & Prifti 2004).[372]

Romani forms a significant element in a variety of in-group slangs such as those discussed in §4.4.2. We can mention here especially Kaliardá, the Greek gay slang recorded by Petropoulos 1993 for which the foreign sources of vocabulary tend to

370 Such grammatical variation as does occur is generally aimed at reducing markers that might aid outsiders in interpretation (cf. Friedman 2011c).

371 See Polenakovikj 1951ab and Jašar-Nasteva 1953abc for additional discussion and bibliography. See also Leschber 2006 on Albanian elements in Macedonian and Bulgarian secret languages. See Sikimić 1992 on the Romanian contribution, especially to secret languages in Serbia and Kaymaz 2003, with bibliography, for secret languages in Turkey.

372 Although Shkurtaj identifies *dom* in Purisht (see below) as being from Russian, in fact the word remained colloquial in Macedonian in the adverb *doma* 'at home.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

be Italian, French, and English as well as Turkish (Vunčev 2017: 47). However, as Asenova 2017 points out, the Turkish elements are basically those that have survived in colloquial Greek despite being excluded from formal registers, e.g., Kaliardá τζοβαΐρι 'jewelry' from colloquial MedGrk τζοβαΐρι 'precious stone' (< Trk *cevahir*) vs. literary πολύτιμος. Asenova 2017 also notes the fact that Romani supplies an important component of productive elements (she identifies about a dozen), e.g., λατσό 'good, beautiful' from Rmi *lačho* 'idem,' as in the Kaliardá λατσολιγγα 'Katharevousa,' where the second element is from Italian. In Stojkov 1968: 226–247, 1993: 340–362, there is a good survey of the topic for Bulgarian, with references, and he makes the point that Romani elements are found in slangs and jargons throughout Europe (and, we can add, the Western Hemisphere) – a point also made by Matras 2002: 249–250 in his discussion of the covert prestige of Romani (with references to Kostov 1956, 1970; Graur 1934; and Julliand 1952, among others, for Romani) – but relative degrees of such vocabulary have yet to be studied.

Turkish slang or informal usage is sometimes taken over into Balkan languages with the same meaning, e.g., Blg *čaktisvam* 'I understand' from Turkish *çak-* (1SG past *çaktım*), lit., 'get, grab' but with the same semantics as colloquial English *I get it* meaning 'I understand it,' or *tarikat* 'clever, cool' (Trk *tarikat* 'dervish order,' cf. Leschber 2007). As Shkurtaj 2004 points out, one of the reasons Purisht, a secret language used by labor migrants in a group of villages in southern Albania, is now moribund is that men going on labor migration more recently usually go to Greece and can simply use Albanian as their secret language.

In the case of Jewish languages, either Judezmo or Yiddish could be the source of Balkan Slavic slang *aver* 'friend' (cf. Heb *ḥaver* 'friend, comrade'), but Yiddish must be the source of Blg *redim* 'I speak/say' (Yid *redn* 'to speak, talk'), which points to a nineteenth-century origin for the term. In Judezmo itself, Hebrew had the high status of a holy language, but it was also available for cryptolectal purposes in secular contexts. Thus, for example Bunis 2011: 32 gives the following examples: *No diburees, ke yodéah lashón!* 'Don't speak because he knows the language' (cf. Angloromani *mursh akai!* '[There is a] man there!' [Matras 2010: 120]) and *los enáim en las yadaim* 'eyes on the hands' (caution against potential shoplifters). In these examples, only the function words *no, ke, los, las, en* are of Spanish origin, whereas verbs and nouns are all Hebrew, but not used in everyday Judezmo. Hebrew for Judezmo thus represents a register that can be both elevated and cryptolectal.[373]

Religion also provided a secret language among the Orthodox Balkan Slavic speakers, albeit only for the clergy. According to Popovski 1951, this secret language involved simply spelling words, but using the Church Slavonic names for the consonants with reduplication of the initial consonant for a vowel, e.g., the

373 The elevated status of Hebrew in Jewish languages serves as the motivation of a joke in which a German in nineteenth-century eastern Poland is accused of a crime and brought to court. Since no one speaks German, a Jew is brought in to interpret. The Jew, finding himself in such an official environment, uses his most elevated Yiddish, which is so full of Hebrew that the German cannot understand the "translation."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

name *Stale* becomes *slovo tvrdo-ta ludi-le.* While the illiteracy of the general population contributed to the code's efficacy, the children's secret language *King Tut* used in the United States follows a similar principle and is quite effective despite the literacy of the general population.

Here we can also mention an important aspect of cryptoglossia observed by Karastojčeva 2001/2002, namely syllabic play, which is an important element in secret languages everywhere.[374] When spoken fluently, such languages are difficult to understand even for native speakers of the base languages, but they are completely opaque to nonnatives, even those who have an otherwise excellent command of the language. According to Karastojčeva, such cryptoglossia is used only by children in eastern Bulgaria, much as only children use such languages in anglophone North America and much of North Macedonia. By contrast such word deformations form an important part of the secret languages of western Bulgaria (and, we can add, parts of North Macedonia). One lesson to be drawn from this difference is, perhaps, that in the more complexly multilingual environment of the western Balkans, the degree of multilingualism was such that word deformation in adult secret languages can be considered as a symptom of societal multilingualism.

Jargons differ from secret languages in that their vocabulary is specific to the profession, hobby, or other occupation they serve, as opposed to secret languages, which routinely have basic vocabulary (eat, drink, man, woman, etc.) in addition to possible specialized vocabulary. The border between jargon and slang is not rigid. Members of a given profession may also have professional slang, as in the case of the Turkish *Karagöz* players cited in §4.4.2. Jargon, as we understand it here, refers to the vocabulary used by members of a community defined by occupation, *sensu lato*, in referring to items defined as pertaining specifically to the occupation. This understanding of *jargon* suffices for our purposes here. It is, in a sense, technical vocabulary.

As an example, we cite the color terminology used in the hobby of dove-raising (*golubarstvo*) in Macedonian, where Turkish terms are used for the names of different types of birds, e.g., *ak kuruk* 'white tail,' *kara kuruk* 'black tail,' *beaz* (Trk *beyaz*) '[pure] white,' *sija* (Trk *siyah*) '[pure] black' (Cvetkovski 2017). This terminology can be treated as a technical subset of the colloquial standard (cf. Friedman 2011d). At the same time, it utilizes two layers of color vocabulary in Turkish itself, native and Arabo-Persian. Jašar-Nasteva 2001: 48 also gives Turkish terminology for horse colors in Macedonian: *abraš* 'horse with white spots,' *dorija (doru)* 'dark-red, brown horse,' *alčo, alatest* 'red horse.' The same types of

374 Thus, for example in North America, secret languages such as the widespread Igpay Atinlay (Pig Latin) as well as less widely known Abinglabish (English, with *ab* added before each vowel) and Opish (where *op* is added after consonants, e.g., *Ingoplopishop, = English*) are exactly analogous to such Balkan children's languages as those recorded for Macedonian by Cepenkov 1972b, e.g., *Štrkolski* 'Storkish' (also called *Gaskarski* 'Goosish' or *Lastovički* 'Swallowish'), where the syllable *grV* or *glV* is added after vowels, so that *vigridogrov* = *vidov* 'I saw.'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

Turkisms occur also in Albanian for color usage in dove- or horse-breeding, e.g., *kara* for a black creature.

A different domain of color terminology is found in relation to animal husbandry. Here we move from contact with Turkish, to contact between Albanian, Balkan Romance, Balkan Slavic, and Greek in a domain that is quintessentially rural, and by its very nature indifferent to the state insofar as is possible. As an example, we take some recent terms that involve white coloring on small cattle, Sobolev 2009a being a useful source for such terminology. In Kastelli (Peloponnesos, Greek), the Slavonicisms *mb'elo* and *mb'ela* are used for 'white ram or lamb' and 'white ewe,' respectively (Map 56), the Latinism *fλ'oro, fλ'ora* for 'white goat' (Map 86), and the Albanianisms *λ'ara* or *mb'artsa* for 'white-bellied goat' (Map 88). This second root turns up in Aromanian *bardzi* 'white-bellied sheep' (Map 56).[375] Another Slavonicism is derived from Slavic *рьрčь* 'billy goat,' which occurs thoughout South Slavic with the appropriate reflex for vocalic /r/. It also occurs in Alb *përçak* (Leshnja, Tosk), Aro *părču* (Turia [Grk Krania], Pindus, Greece), and Grk *purčus* (Eratyra, in western Greek Macedonia) (Sobolev 2009a: 172). See Kahl 2007 for a study of contact among Albanian-, Aromanian-, and Greek-speaking shepherds.[376]

Finally, we can mention what is probably the best-documented Balkan contribution to occupationally based vocabulary, namely the role of Italian and Greek in shaping Turkish nautical terminology, and by extension, mariners' jargon more generally in the Mediterranean, and thus the Balkans. Kahane et al. 1958 meticulously documents 878 words, 154 of them Greek, that passed from Italian and Greek into Turkish from the thirteenth through the eighteenth centuries, though in the course of so doing, numerous words are documented for Greek nautical usage itself, and for other Balkan languages, with various directions of diffusion. A few illustrative terms are Itl *boma* 'boom,' the source of Trk *bomba*/*bumba* 'spanker boom,' but also Grk μπούμα and Alb *bumë* 'boom'; Vtn *flama* 'pennant,' source of Trk *flama*/*filama* 'streamer,' and related to Grk φλαμούρο(ν) 'nautical banner,' an alteration of Byzantine Grk φλαμοῦλον, from the derivative Lat *flammula*, itself the source of Alb *flamur* 'flag' and Rmn *flamură* 'pennant'; Grk παξιμάδι(ν) 'biscuit,' source of Trk *peksimat*/*beksimet* (and variants) 'hard biscuit,' but also Alb *paksimadh*/*peksimat* (and variants), Rmn *paximat*/*pesmet* (and variants); and words for 'harbor,' Byzantine Grk λιμένιον, the source of OSrb *limenь*, and Trk *limen*, a variant of which, *liman*, is the source for the word in other Balkan languages, e.g., Alb *liman*, BSl *liman*, later Srb *limān*, Rmn *liman*, and ModGrk λιμάνι (thus, a reborrowing). See now also Panzac 2008 and Nolan 2020 and the references therein.

375 Mac *barz* can also refer to a 'person with salt-and-pepper hair.'

376 See also Friedman 2011d for more on the Balkan features of basic color terms (black, white, red, but also green).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006

### 4.4.4 Other Sources and Types of Register Differences

Among the other sources of differences in register, we can identify shared ideology, shared experiences of extra-Balkan influences, and shared localizations. Shared ideologies that valorized earlier stages of a given language as more "pure" or "uncorrupted" were most powerful in Greek, where for much of the nineteenth and twentieth centuries, a puristic, Atticized, consciously archaizing, colloquial called *Katharevousa* dominated formal registers and was exemplary in the theorization of the term *diglossia* by Ferguson 1959. The relationship of Latin to Romanian and of Church Slavonic to Balkan Slavic was strong in this respect but never split the new standard languages in two as happened with Greek. For Romanian, French and Italian were more modern related languages of status and power, for Bulgarian it was Russian and for Macedonian, Serbian, especially after the Tito-Stalin break of 1948.[377] Attempts at Sanskritizing Romani have so far not had significant results. In terms of twentieth century power relations, French, German, and Italian have all had effects on their spheres of influence, e.g., Italian in the Albanian of Albania whereas in the Albanian of former Yugoslavia German supplied parallel vocabulary, e.g., *skapamento* versus *auspuh* 'muffler' (see also §4.2.1.7). Today, as almost everywhere else, English is a major source of new vocabulary. As we noted in §4.2.1, English in the Balkans is the Turkish of the twenty-first century. Moreover, English has entered every level of vocabulary: technical, unmarked colloquial, slang, etc. The one other source that we can note here that is specifically Balkan is localized language contact. Thus, for example, Albanian *çupa* 'girl' (DEF) and *bishka* 'pig' (DEF) are the source of these same words in southwestern Macedonian dialects. Today's Macedonian speakers are often unaware of such words' Albanian origins and consider them their own, emblematic dialectisms. Such examples could be cited everywhere in the Balkans where older patterns of dialect contact have been superseded by new borders.

377 However, even in 1944, Blaže Koneski opposed those who wanted to bring in Russian linguists to standardize Macedonian (see Friedman 1993b).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 05:14:35, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781139019095.006