5 Phonology

5.0 Introduction

The most striking facts about Balkan languages that attracted the early attention of linguists centered on aspects of grammar, where grammar subsumed what would today be called morphology and syntax; note Kopitar's observation cited in §2.2.1 concerning "grammar" ("Sprach**form**"). Nonetheless, as indicated in §4.1, much of the early attention focused on the lexicon. Yet it also became apparent that there was more to the convergence area than just grammar and lexicon, and in particular that the phonological component of these languages was also relevant.[1] In this chapter we examine the evidence for phonological convergences, where "phonological" is understood in a general nontechnical sense of 'having to do with sound.'

5.1 Some Background and Prospects

As discussed in §2.2.3, fully half of the Balkan common features taken as diagnostic for the sprachbund in Miklosich 1862 are phonological. Among these were the change of *n* to *r* ("rhotacism") in Albanian and Romanian, the prevalence of schwa, and the occurrence of nasal onsets with voiced stops.[2] Similarly, as noted in §2.3.2.1, several putative phonological convergences were stated in Selišč ev 1918, 1925.[3]

1 It may be that the many obvious convergent points of grammar and lexicon led some scholars to overlook the possibility of phonological convergences, both in the Balkans and perhaps more generally. The appearance of a substantial handbook treatment of "areal sound patterns" (Blevins 2017) is a modern approach to contact phonology and represents a significant advance over Jakobson 1931a/1962.

2 See also §3.2.7, for discussion of these facts from a methodological standpoint. It is now clear that the areal nature of each of Miklosich's "phonological" features ranges from irrelevant to in need of significant nuancing.

3 Jannaris 1897: 124–125 recognizes that phonologically northern Greek (citing Velvendos, northeast of Kozáni at the southern edge of Greek Macedonia) differs significantly from southern Greek (citing Crete). He attributes this difference to northern Greek dialects having been influenced by the other Balkan languages, and he claims that the southern dialects have "withstood foreign influence." Jannaris goes on to promote the ideology that therefore, for the neo-Hellenist, only southern Greek is worth examining. While Jannaris is correct that language contact certainly has a role in differentiating northern and southern Greek, his formulation represents a classic example of Irvine & Gal's 2000 *erasure* and *fractal recursion*. Jannaris' construction of southern Greek as having "withstood" language contact erases the fact that all of Greek has participated in many contact-induced Balkan linguistic changes, as well as the fact of contact influences in southern Greek itself (cf. Leluda-Voss 2006 or Orfanos 2014). The formulation also represents recursion in that nineteenth-century Western

Sandfeld 1930 did not pay much attention to putative sound-based concordant features. The bulk of his book falls into his large chapters: II on loanwords and IV on grammatical features (totaling 137 of 201 pages of text, excluding the introduction). He discussed shared aspects of phonology in various pairs of Balkan languages interspersed amongst grammatical features in his chapter III on "*Concordances entre différentes langues balkaniques en dehors du lexique*" ('Convergences among different Balkan languages outside of the lexicon'). Sandfeld draws attention to a considerable number of such convergences involving sounds, and many of these have been repeated and elaborated on, and their number has been increased in subsequent treatments of the Balkan sprachbund, particularly as scholars have sought to uncover and develop wide-ranging features that parallel the morphosyntactic phenomena of broad distribution in the region. Two that are often mentioned in the literature, albeit not in Sandfeld's work, are given in (5.1):[4]

(5.1) a. the presence of a (stressed) mid-to-high central (thus, schwa-like) vowel
b. the presence of *i-e-a-o-u* in the vowel inventory without phonological contrasts in quantity, openness, or nasalization

Other researchers have compiled even larger lists, as becomes evident below. Still, the clear emphasis in Balkan linguistics has been on domains other than phonology.[5] To our knowledge, there are relatively few monograph-length treatments of Balkan phonology.[6] In addition to Afendras 1968 and Sawicka 1997, there are Sawicka 2014, 2023, on Balkan phonology in general, Sawicka & Cychnerska 2018 on Macedonian phonology specifically in its Balkan context, and Sawicka & Dargiel 2018: 35–42, on Balkan syllable structure in the context of Albanian syllable structure.

After due consideration, we reject all of the various heretofore proposed pan-Balkan phonological features. However, in line with the importance we place on localized convergence (see §3.3), several features covering subsets of the languages or regional dialects do turn out to be significant. In this respect, therefore,

sources regarded the Greek of their era as "corrupted" by foreign influences (cf. Herzfeld 1982). Jannaris applies this ideology recursively within Greek, so that the south is less "corrupted" than the north.

4 To the best of our knowledge, Miklosich 1862 is the source of (1a) about the schwa-like vowel (Naylor 1980: 58 explicitly says it is) and Havránek 1933 is the source of (1b) about the vowel system. These two features are discussed or at least mentioned in virtually all handbook-like treatments of the Balkans (Schaller 1975; Banfi 1985; Feuillet 1986; Asenova 2002; Sh. Demiraj 2004; Steinke & Vraciu 1999) and smaller-scale encyclopedia-type surveys (Joseph 1992a, 2020a; Hinrichs 1999b; and Friedman 2006a, 2021a).

5 In this way, the Balkans differ from linguistic areas such as the Caucasus, the Northwest Coast of North America, and South Asia, where phonological features such as glottalization and retroflexion are among the most salient commonalities discussed in the literature (see Hock 1975, for instance, on retroflexion in South Asia, and Tuite 1999 on the Caucasus).

6 There are of course treatments of the phonetics of some of the individual languages, e.g., Adamou & Arvaniti 2014, some of which (e.g., Minissi et al. 1982 on Macedonian) contain some comparative observations concerning Balkan phonetics. Nonetheless, phonology is generally neglected in Balkan linguistic studies. As Table 4.1 in §4.1 shows, in nearly every Balkan linguistic handbook considerably less space is given over to phonology than to any other domain, with no work devoting more than twenty-two pages to the topic, and all showing an average of twelve pages on Balkan sounds and sound structure, as opposed to 19.4 pages for the lexicon and 92.8 for morphosyntax.

our findings support Sandfeld's original 1930 account. Despite our skepticism about wide-ranging phonological convergences, the various proposed features merit attention, albeit critical. Thus, while these phenomena must be discussed in the context of Balkan linguistics, in the end it is only the local convergences that are Balkanologically significant.

We therefore give full coverage to putative phonological convergences found in the literature. To frame the discussion, we first justify our localistic approach in phonology and then give selected illustrative case studies. We next offer extensive consideration of these features, working through them segmentally by "vowel-related versus consonant-related," crossed by the distributional parameter of "(more or less) pan-Balkan versus localized." This latter parameter recalls somewhat imperfectly the organizational scheme in Sandfeld 1930, where the abovementioned chapter III on "*Concordances entre différentes langues balkaniques . . .* " ('Convergences between different Balkan languages . . . ') treats parallels found in different pairings of the languages, and chapter IV on "*Concordances générales en dehors du lexique*" ('General convergences outside of the lexicon') treats more widely distributed parallels. In each case, we further test the degree to which a feature has legitimacy as a Balkanism in order to evaluate claims concerning Balkan sprachbund phonology. This organization leads to some features being discussed in more than one place, but it offers maximal coverage of the important material. Finally we turn to other aspects of the phonology of these languages, including prosody, morphophonemics, and sound-based expressivity.

5.2 Localized Phonological Convergence and Bilingualism: The Mechanism of Phonological Borrowing

As already noted, the valid features here occur in subsets of the Balkan languages, not in the region as a whole. Our contention is that attempting to find sound-based Balkanisms that apply to the entire region completely misses the point about the formative processes and circumstances that have led to the Balkan sprachbund and especially to the respective phonologies of the languages involved. Rather, for phonology, it is a localized, cluster-based approach to the sprachbund, examining situations on a local level instead of large areas on a macro-level, that is revealing (see §3.3).[7] This approach was first enunciated by Seliščev 1925: 49 and elaborated on by Hamp 1979, 1989a, and Friedman 2008b. Ultimately, although lexical borrowing is in part responsible for the spread of phonological features, the most critical force underlying these local convergences is bilingualism as described

7 Our reference here to macro- and local levels, while perhaps reminiscent of the distinction in Hinrichs 1999b: 431 between "Makrobalkanismen" and "Mikrobalkanismen," is actually quite different. Hinrichs distinguishes between large-scale overarching structural properties, such as a tendency towards analytic expression, and highly particularized lexically/constructionally specific features, such as the syntax of a single word, whereas for us the relevant notions are to be interpreted areally (i.e., geographically).

in §3.2.1.1, i.e., *some* working knowledge of another language on the part of individual speakers and groups of speakers. Intensive use by individual speakers of co-territorial languages is especially important. As described in §3.2.1.1, such a situation often leads a substratum effect involving transference as the pronunciation habits of one's native linguistic system carry over into the other secondarily acquired linguistic system(s). As Sawicka 1997: 16 notes, "conservation or loss of given sounds is evidently determined by the sound system of a coexisting dialect."

Thus, for example, Hamp 1989a notes the parallel accentual pattern of Albanian and Romanian, observing that in each language stress falls on the final syllable of the stem of a given lexeme. That is, in general, the stress shifts rightward with the addition of a derivational suffix, e.g., the agentive suffix *-tor-* in Albanian or the deverbal adjectival suffix *-tor-* in Romanian, but not with the addition of an inflectional ending, such as the Albanian dative plural ending *-eve* or the Romanian definite dative singular ending *-lui* or the feminine singular ending *-e*; see (5.2/5.3):[8]

(5.2) Albanian
púnë 'work.NOM.SG / punëtór 'worker.NOM.SG' / punëtóreve (DAT.PL)
katúnd 'village.NOM.SG' / katúndeve 'villages.DAT.PL'/ katundák 'rustic', katundarí 'peasantry' vs. katundári 'villager/peasant.NOM.SG.DEF '

(5.3) Romanian
a folosí 'to use' / folositór 'useful.M.SG'
folositór 'useful.M.SG' / folositoáre (F.SG)
cîíne 'dog.NOM.SG' / cîínelui 'to the dog.DEF.DAT.SG'

Hamp's interpretation of such facts (1989a: 47) is that "historically Romanian is Latin spoken with an Albanian stress system," that is, the "Danubian Late Latin" of Dacia was filtered through the grammars (that is to say, speech habits) of one group of speakers of Albanoid or a closely related language. Similarly, Petrovici 1957: 43, based in part on the parallel pre-iotacizing development of word-initial *#e > je-* in Romanian and Slavic (e.g., Romanian *eu* 'I,' phonetically [jɛw], from Latin *ego*), comments that Romanian is "*einer romanischen Sprache in slavischer Aussprache*" ('a Romance language with a Slavic pronunciation'). Although this general view has recently been criticized by Petrucci 1999: 43–49 as not being sufficiently based on the facts of Balkan Slavic phonology to warrant so broad a conclusion for some aspects of Romanian pronunciation (see §5.6 on palatalization), he does not extend his critique to the pre-iotacization phenomenon (on which see §5.4.3.5 below).[9]

In §3.2.1.3, we argue that reverse interference can occur when the secondarily acquired language affects aspects of a speaker's native language. This reverse

8 There are some exceptions to this general pattern; for instance, the Albanian adverbial suffix *–azi* forms derived manner adverbs from adjectives (including participles) but does not attract the stress, as in *fshéhurazi* 'secretly,' from *fshéhur* 'hidden, secret' (participle of *fshéh* 'hide').

9 In fact, nonlinguist anglophones who know Bulgarian and Italian endorse Petrovici's impression upon first hearing Romanian (so VAF, for instance).

interference proves essential to understanding various cases of Balkan phonological contact. Such effects can be hard to demonstrate conclusively, as discussed below, without knowing definitively which language is the first language and which the second language and the extent of use of each, but as a type of contact effect it must always be entertained as a possibility. A few cases from Albanian dialects illustrate the utility of recognizing reverse interference in understanding Balkan phonology.

Thus, for example, Sandfeld 1930: 104 states that in Arvanitika, one finds, as he indicates it, *mnj* for what other Balkan Tosk Albanian dialects have as *mj*, e.g., *mnjekrë* 'chin; beard' (vs. general Tosk *mjekër*) and he notes that this shift of *mj* to *mnj* is "*comme en grec*" ('as in Greek').[10] He is referring to the phenomenon, noted e.g., by Thumb 1912: 23, §30, whereby, for instance, μιά 'one.F' (via rightward stress shift from earlier μία) is pronounced with a palatal nasal [ɲ], thus [mɲa], where the [ɲ] has arisen out of the *-mj-* sequence (with [j] from unstressed [i] before a vowel). Although Sandfeld uses a transcription that is phonetically unclear, he most likely is indicating a pronunciation involving nasality and palatality, thus something like the [ɲ] of Greek. We can thus go further and state that this example shows imposition of Greek pronunciation habits onto Arvanitika. That is, if Greek is a second language for Arvanitika speakers in Greece – and in some cases even the dominant language given its social importance (and official dominance) in all Arvanitika-speaking parts of Greece – the *mnj* (i.e., [mɲ]) pronunciation would be a case of reverse interference from Greek as a second language onto Arvanitika as a first language (i.e., first-learned, mother tongue) for these speakers.[11] However, we should also note that the shift of *mj* > *mnj* occurs elsewhere in Albanian, e.g., facultatively in the East Central Geg dialects of Macukull and Tanushaj (Mac Tanuše), and consistently in the Arbëresh of Karfici in Calabria (Gjinari 2007: 104). In the case of Arbëresh, contact with Griko might be contemplated, but is probably not an issue, since stressed *-í-* is preserved in Southern Italy Greek (Thumb 1912: 12, §10.1) so that μία would remain as such. In the case of Macukull and Tanushaj, it is worth noting that the two points between these villages, Gur-Lurë and Suhodoll, appear to be the nucleus for the change and have carried it further, with *nj* for *mj* consistently. However this is an isolated innovation in the overall context of Albanian (see also footnote 116).[12]

Similarly, Montenegrin dialects such as Mrkovići (Stevanović 1935: 42–43) and Crmnički (Miletić 1940: 279–282) have velar /ɬ/ as in Albanian, and Albanian clear /l/, rather than the palatal /l/ of Bosnian-Croatian-Montenegrin-Serbian (BCMS) (cf. also Morozova & Rusakov 2018a, 2018b; Morozova 2021),

10 We are using here Sandfeld's notation.

11 Thumb writes μνιά in discussing the pronunciation of μιά; it is not clear, nor does it really matter for the issue at hand of Arvanitika pronunciation, whether he intended [mɲja] or simply [mɲa]. The same can be said about Sandfeld's *mnj* for Arvanitika. The point is that the Greek and Arvanitika pronunciations converge, with Greek a possible causal factor for the Arvanitika.

12 Gjinari 2007: 104 does not record *mj* > *mnj* for the two Arvanitika points, so either Sandfeld's observation about Arvanitika was quite localized, or the data in Gjinari 2007 do not reflect the many Arvanitika-speaking villages of the past.

presumably via reverse interference from Albanian into the local BCMS. Hamp 2010 and Dombrowski 2013: 149–153 discuss liquid realizations in BCMS and Albanian, and on Balkan laterals more generally, see §5.4.4.8.

Also, Hamp 1989b: 203 gives a particularly telling set of shifts in Geg Albanian that appear to be reverse interference from Slavic into Albanian, since Albanian is fixed, in a sense, and the variable is the local Slavic idiom; he describes the situation as follows:

> Much of Geg (including Gusî [in Montenegro, VAF/BDJ]) shares a consonantal characteristic with the neighboring Slavic languages. In Northern Geg *ḱ* and *ǵ* are articulated as affricates in exactly the same fashion as Serbo-Croatian *ć* and *ʒ'* (orthographic *đ* or *dj*); in Dukagjin in northern Albania, the articulation shifts further to *ś* and *ź*. In Kosovo these merge, in Albanian and Serbo-Croatian, with *č* and *ǯ* (orthographic *dž*), and Makedonski [i.e., modern Macedonian, VAF/BDJ] has of these pairs only *č* and *dž*; however, Makedonski possesses also *ḱ* and *ǵ* [which correspond etymologically to Serbo-Croatian *ć* and *ʒ'*, VAF/BDJ]. Tetovo [Albanian of North Macedonia] shows the following interference innovations: **ḱ ǵ* > * *ć ʒ'* > *č ǯ*, and **tj dj* > *ḱ ǵ*, thereby exactly matching Makedonski in distinctive feature structure.[13]

A cautionary note must be sounded here, though. These cases can be interpreted as showing reverse interference or simple convergence, and there are well-established results (see §3.2.1.3) in controlled cases where there was no doubt about which language was the first language for the speakers in question. As for Arvanitika and Greek, however, it is conceivable that, if more were known about the early days of Arvanitika in Greece (late Medieval period), it could turn out that later speakers of Arvanitika were originally native speakers of Greek who shifted to Arvanitika; in such a scenario, the apparent reverse interference effects would instead be a substratum transfer effect such as that posited for the stress placement facts of (5.2) and (5.3) above. This is not to deny reverse interference as a genuine phonological contact phenomenon or to suggest that we have particular knowledge about the social interaction of Arvanitika speakers and Greek (or Slavic) speakers in that period, but only to say that the phonological effects in contact situations involving bilingualism can be sufficiently complex as to require one to keep an open mind in determining what happened in the past.

A similar reservation applies to Albanian–Slavic interaction. Given that these developments occurred after the Slavs reached the Balkans but in the absence of adequate documentation, we cannot be sure of the direction of interference. However, since the biphonemic reflex of Common Slavic **tj* /*dj* extended further north into central North Macedonia, probably until the twelfth century or so (when it was incorporated into the Serbian Empire), we must posit a shift in pronunciation among the local Slavic-speaking population, at least in what is now central North Macedonia, as evidenced by *gakji* 'britches' but *gašnik* (< **gaštnik*) 'belt holding

13 In fact, the development of *tj/dj* in East Central Geg to mellow dorso-palatals like those of Macedonian is found in all the Geg dialects of North Macedonia except Zajaz in the Kičevo (Alb Kërçova) region (Jusufi 2011: 183).

up britches' in some central Macedonian dialects. It is striking, however, that the Albanian dialect of Zajaz, which is in the same region where there is a *gakji/gašnik* opposition in Macedonian, is unique in Macedonian Geg in not having the development of jotated dentals into palatal stops (see footnote 13). It may well be then, that as the monophonemic speech habit spread southward, it affected both Slavic and Albanian.

To elaborate on the role of bilingualism, we stress that besides interference (reverse and otherwise) and a substratum effect, there are two other dimensions to consider in the phonology of a bi- or multilingual speaker: simplification in second language acquisition and demarcative expressivity in a multilingual situation. Both are potentially huge topics and the former is treated above in Chapter 3, but for our purposes here it suffices to give some essentials that contribute to an understanding of the Balkan phonological situation.

To address simplification first, as discussed in §3.2.1.5, one way in which simplification can be manifested is by speakers of one language accommodating in the direction of the particular simplifications that L2 speakers of that language make. What would result is a compromise system, characterizable as a kind of lowest common denominator phonology.[14] This system would be shaped by a sort of mutual simplification that alters the target language both as an L2 variety and as used by native speakers.[15] A Balkan example is the clear vowel system and especially the absence of nasalization as an "overlay" feature, as discussed below in §5.4.1.6.

As for the dimension of expressivity and social boundary marking in multilingual contexts, we note first that bi- or multilinguals have choices regarding language use that monolinguals do not. This fact underlies the important insight of Weinreich 1968 (see §4.3.1.10, footnote 124 for details) that multilingualism increases an individual's range of expressivity, for the simple reason that there is more material to draw on for expressive purposes. Furthermore, that material necessarily – and literally – is *exotic*, standing outside (*exo-*) of the other language's system; sounds from the second system would thus be able to contribute to that expressivity.[16] Similarly, recognition of certain sounds as being outside of a system allows speakers to use the sounds as elements that define the limits of one language and establish linguistic boundaries; by the same token, sounds can mark a word as

14 This term comes from A. Schmidt 1985: 146 who sees accommodation as the causal mechanism in the Dyirbal contact-affected phonology: "the norm of each in-group [studied] is similar to the careful Dyirbal style of the least-fluent member (i.e., a 'lowest common denominator effect'). This suggests an interlocutor rule that speakers of the in-group modify their Dyirbal to a level that all members can respond in. The norm must be within the competence of all peer-group members."

15 This is sometimes referred to as *foreigner talk* (on which see Ferguson 1971).

16 Herbert 1990abc discusses the entry of clicks into the Nguni group of Bantu languages in southern Africa, and especially their occurrence in an avoidance language, a special register, known as *hlonipha*. He advocates the position, following Faye 1923–1925, that through contact with Khoisan speakers who had clicks in their languages, Bantu speakers were provided with a suitably expressive set of sounds, various clicks, that they could exploit in this special register, and that from that register the sounds entered more ordinary registers (cf. §3.4.1.1).

being within or outside of a given system. In this sense, they can reflect ideology about what sorts of elements belong to one system or the other.

Thus, we argue that phonological Balkanisms, to the extent that they exist, come about through the entry of features of pronunciation into a language via bilinguals, where "bilingual" has the fairly broad definition of §3.2.1.1 that includes speakers with an imperfect command of L2. In particular, these speakers either draw on the resources of the second language and bring new, "exotic" sounds into their native language or else they impose their native phonology on the second language and alter the second language when they use it. Additionally, they may even undergo the effects of reverse interference on their native language from their second language. We posit further that bilinguals use and exploit their full range of registers for various purposes, including the social marking and delimiting of boundaries. These features can then spread into the usage of monolinguals in either language modeling their pronunciation on that of the bilinguals in a given community, i.e., accommodating to them. Alternatively, or additionally, particularly in the case of transference effects from the native language projected onto a second language, it may be that several different individuals essentially arrive at the same alteration of the second language; this is understandable if the second language is being filtered through the phonology of speakers of the same native language, since the same systemic elements present in one speaker of a given language would be present in another speaker of that same language.[17]

Lexical borrowing can also be the point of entry for novel phonology, especially for nonfully bilingual speakers in the speech community. However, as emphasized below in §5.3, in which several representative case studies are presented, including an extensive one involving Aromanian, neither lexical borrowing alone, nor the filtering function of one's native phonology alone, can explain all the relevant facts. In some instances, the degree of bilingualism and familiarity with another language must be factored in. Moreover, the social status of the language can be a factor, as can individual speakers' attitudes about how foreign words should be pronounced, an example being the conscious imitation of Turkish phonology in the pronunciation of Turkish words in several Balkan languages and the preservation of voiceless aspirates as a distinctive feature in Romani, showing overt resistance to foreign influence.

The role of lexical borrowing in the introduction and spread of foreign phonology is emphasized by Stankiewicz 2002: 369 in his discussion of parallel phonological features and processes found in Albanian and Romanian; whatever the merits of this particular list of features for these

17 This sort of independent but parallel development has not had the attention in the literature on language change that it should, but cf. Janda & Joseph 2003: 83 regarding some such lexical changes in American English slang. Similarly, Thomas 2002: 179 speculates that four speakers he worked with who had similar but nonidentical features show "projections of individual identity," suggesting that even the similarities could have arisen independently in each speaker.

particular languages (see below in §5.4.4.10 for further discussion), Stankiewicz's invocation of borrowing brings to light this important conduit for phonological innovation and diffusion:

> The linguistic exchange between pre-Romanian and pre-Albanian, which must have begun centuries before the arrival of the Slavs, *i.e.,* when the two languages were still contiguous, provides a lucid example of a process that was to repeat itself throughout the linguistic history of the Balkans, a process in which extensive lexical borrowing was invariably followed by the diffusion of phonological and morphological traits. The phonological diffusion is best illustrated by the series of changes that affected, to a similar extent, the phonology of both languages. They included the change of Latin short *a* to Rom[anian] *ă*/Alb *ë* (= *ə*); the narrowing of the sequence *en* (followed by a dental) to *in*, and of *an* to Rom *în* and Alb *ën*; the shift of the clusters *kt*, *ks* to *pt*, *ps* in Romanian, and to *ft*, *fš* in Albanian; the change of the velars to strident palatal affricates (before an original short *-i*) in Romanian, and to palatal stops or affricates in Albanian; the palatalization of the sonorants *ll*, *l*, *n* and their subsequent change to *j*; perhaps, the shift of intervocalic *-n-* to *-r-*. The developments in question did not take place at the same time nor without some differences. Thus, the rhotacism of *-n-* eluded most of the Romanian dialects as well as Albanian Geg; the Albanian *ṅ* did not change in all dialects to *j*, while Romanian did not follow Albanian in palatalizing *r* and changing it to *j*. The diffusion of phonological features did not and could not exclude diversity.

Once the borrowing of individual lexical items enters the picture, two further concepts concerning loanword phonology become relevant, echoing in part what has already been said. First is the matter of phonological nativization – or the lack thereof – (so-called "adaptation" versus "adoption," as in Hock 1991: 408). What is at issue here is the extent to which speakers of the borrowing language either simply take over loanwords in their foreign form (adoption) or instead re-cast them into their native language's phonological patterns (adaptation); see Hock 1991 for an extensive discussion with copious examples. We emphasize here that the adoption versus adaptation strategies are not structurally determined by the nature of the borrower's phonological system, as is often claimed; rather, we see the borrower as having some degree of choice as to whether to sound like a foreigner or not (see §3.2.2.10, footnote 122, and compare the notion of *authentication* of foreign sounds on the part of a borrower discussed by Matras 2009: 225, 228). Second, the claim has been made, e.g., by Thomason & Kaufman 1988: 74–75 and reformulated slightly in Thomason 2001: 70–71, that a particular level of borrowing intensity – level 4 ("strong cultural pressure: moderate structural borrowing") on the 1988 scale of 5, or level 3, involving "more intense contact," on the 2001 scale of 4 – is needed for the introduction of new phonemes into native vocabulary. This means that not all contact and borrowing situations offer the social milieu in which foreign sounds, having once entered a language via lexical borrowing, can diffuse into indigenous lexical items. The spread of foreign phones from (adopted) borrowed items into native words so as to create new phonological contrasts in native words requires sprachbund-like conditions, i.e., intense and sustained contact with significant bilingualism.

To return to bilingualism per se as the force behind phonological convergence, an important aspect of societal bilingualism allows for an explanation of the localistic effect noted above for phonological diffusion. Bilingualism is essentially a local phenomenon, determined by the languages spoken in a given, sometimes quite small, area; it would follow then from the key role of bilingualism in phonological convergence that any sound convergences would be found on a local scale. Under this view there are no phonological Balkanisms per se except for the highly localized ones that are found in smaller clusters of the languages or in local varieties of the languages in contact with one another. Another way of stating this key notion is that there is no Balkan phonology but rather there are only Balkan phonologies, as Friedman 2008a puts it. With Balkan convergent phenomena discernible only by focusing localistically on the dialect level, there will necessarily be many phonologies since there are many locales where contact has occurred.

We are thus in complete accord with the views of Sawicka 1997: 9, who astutely observes that "as convergence is caused by bi- or multilingualism, it can also be expected that certain phenomena will not occur on the whole territory of the Balkans, but they will be met in microregions, which, for that matter, often extend beyond the territory of the Sprachbund." Her 1997: 79 further observation is instructive here as well: "As far as phonetics is concerned . . . we cannot speak anymore about a Balkan community."[18] This is a more pessimistic view than is taken in other accounts, even those handbooks that implicitly, through their limited coverage of phonology (see above footnote 6), suggest that there is little in the way of a phonological side to the sprachbund. However, it is reminiscent of the negative stance taken by Ivić 1968, who writes of an "*alliance linguistique*" ('linguistic alliance') phonologically in the Balkans rather than a sprachbund ("*ligue*") per se. Indeed, for Ivić, the most interesting aspect of Balkan phonology was the set of common features one can find between Romanian and Eastern Bulgarian, what he referred to as the "*le noyau de la ligue phonologique*" ('core of the phonological sprachbund') (1968: 140–141). Although geographically opposite (and consistent with other linguistic levels), the phonologically most interesting part of the Balkans for Sawicka 1997: 11 "is located in the South-Western part of the Balkans [inasmuch as] Greek, Albanian and Southern Macedonian dialects share some very special phonetic features and form a second centre of the phonetic Balkan Sprachbund." In both instances, it is clear that these observations are predicated on a localized view of contact-induced phonological convergence.

5.3 Localized Phonological Convergence and Bilingualism: Some Case Studies

The few examples just given suggest the importance of bilingualism in the borrowing of phonology. To elaborate on the actual processes involved, we present

18 The issue of Balkan phonetics vs. Balkan phonology is taken up in more detail below, in §5.4.6.

here a few case studies showing the effects of localized bilingualism, including an extended examination of Aromanian loanword phonology. Aromanian is particularly instructive in this regard since different dialects of Aromanian have been in contact with different second languages, resulting in differential effects in the phonology. Following the discussion of mechanisms of phonological convergence in §5.2, we look to lexical borrowing as the point of entry for phonological diffusion. Moreover, the two further issues – simplification and expressivity – raised by utilizing this pathway for novel phonology are relevant in the ensuing discussion of Aromanian.

At issue are facts from various Aromanian dialects and the ways in which they interact with co-territorial languages, especially the fact that different Aromanian dialects treat loanwords differently. There is no reason to claim that structural factors play any role in determining these differences. It is the social setting for the interaction and the presence of different co-territorial contact languages that in each case constitute the decisive factors.

For instance, it was noted in the early twentieth century (Sandfeld 1930: 103–104) and subsequently (Caragiu-Marioţeanu et al. 1977) that southern Aromanian, for which Greek is the dominant L2, has the Greek fricatives *γ ð θ* in loanwords from Greek, such as those in (5.4); vowel differences between the Greek orthographic form (based on southern dialects) and the Aromanian (e.g., Aro [i] for Grk <ε>) reflect vowel developments found in northern Greek dialects that are the source of the Aromanian borrowings (and see below on northern Aromanian):

(5.4)
θ : θámî ‘miracle’ (< Grk θαύμα)
θimél^{u} ‘foundation’ (< NGrk θιμέλιου / StGrk θεμέλιο)
θar ‘courage’ (< Grk θάρρος)
anáθima ‘curse’ (< Grk ανάθεμα)[19]
ð : ðáskal^{u} ‘teacher’ (< Grk δάσκαλος)
aðínat^{u} ‘powerless’ (< Grk αδύνατος)
ðíspoti ‘bishop’ (< NGrk δισπότς / StGrk δεσπότης (vocative δέσποτε))
γ : aγru ‘wild’ (< Grk άγριος)
γambró, γrambó ‘groom’ (< Grk γαμβρός)

Caragiu-Marioţeanu 1968: 47 (cf. also Saramandu 1984: 430–431) points to a structural explanation for the phonological shape of these loanwords, without stating the matter explicitly. She points out that the occlusives of Aromanian form neat square-like oppositions involving correlations of sonority (voicing) and frication, e.g., for the labials and for the alveolars (so also for prepalatals and palatals); see (5.5):

19 OCS Cyrillic had Greek theta in its alphabet but Church Slavonic has both *anafema* and *anatema* showing adaptation to the absence of *θ*. The Slavic *f*-form generally represents a more learned attempt at adapting the Greek insofar as *f* was also originally nonnative to Slavic. For later South Slavic one always finds *t* for Greek <θ>, whereas in Russian Church Slavonic and Standard Russian, *f* was favored (over colloquial /x/). Thus Grk Θόμα(ς) > Russ *homa/foma* vs. SSl *toma* (the last name of a certain American linguist represents the colloquial Russian version). Rmn *anatemă* is from South Slavic, but a hybrid form *anaftema* is also attested (cf. Sandfeld 1930: 104).

(5.5)
p – f t – s
| | | |
b – v d – z

Further, she notes that *γ ð θ* similarly fit into such patterned squares of phonological oppositions perfectly in the dentals and the velars; see (5.6):

(5.6)
t – θ k – h
| | | |
d – ð g – γ

The implication is that this system was ripe for the borrowing of these phonemes, since the necessary structural oppositions for them were already present. Under such a view, there is a structural reason for *γ ð θ* not being altered in the course of the borrowing of these loanwords from Greek.

While an interesting viewpoint, it leaves some details unaddressed: why, for instance, would *ð θ* oust *z s* in the patterned square of dental oppositions? How would the *z s* opposition fit then into the system? Why would the dentals not then constitute a six-way opposition? And, why would the once triangularly arrayed velar oppositions (*k-g-h*, in the absence of a voiced fricative) shift to a square pattern as in (5.6) by taking in *γ*? One can suppose that the square array in the labials affected the velars, but why so if the dentals either left *z s* out or had an extended six-way array, so that there was not agreement in the array at all points of articulation anyway? We therefore take the adoption without nativization of these sounds instead to have been socially motivated, and more precisely to have resulted from bilingualism on the part of these Aromanian speakers in Greek. That is, speakers' familiarity with Greek, we suggest, was behind their ability to allow Greek phones into their Aromanian without altering the sounds at all.[20]

The need to take social factors such as familiarity into account is found repeatedly in studies of contact phonology inside and outside of the Balkans. Jašar-Nasteva 1970 and Matras 2009: 228, for instance, mention Macedonian Turkish loanwords with /ts/ from Macedonian. Matras does so in connection with his notion of *authentication*, involving speakers who "are able to identify and produce" the sounds of the original, donor language (see also §3.1.2.3 and

20 See, however, Gołąb 1984a: 40, who provides squares for all the obstruents of Kruševo Aromanian: p-b-f-v, t-d-θ-ð ; c-ʒ-s-z, č-ǯ-š-ž, k'-g'- *γ*'-χ', k-g-*γ*- χ. However, in the same work, Gołąb is clear that social factors are involved in whether Greek-type fricatives are used or not. This accords with our view that on theoretical grounds, we would dispute any claim that structural factors are at work, since we feel that social factors can always override structural ones. For instance, the loss of the *rr*/*r* distinction, in favor of the *r*, as discussed below (§5.4.4.9.3), could be seen as the loss of the marked member of the opposition, but that leaves the interesting correlation unexplained whereby it is lost in precisely those areas where co-territorial dominant languages lack the distinction, such as Turkish in Albanian or Judezmo towns, as opposed to rural areas, in the Ottoman period.

§3.2.2.10, footnote 122). Moreover, Thomason & Kaufman 1988: 32–33 give an example from Siberian Yupik in which loanwords from Russian in the nineteenth century, when familiarity with Russian was rather rare among these speakers, were adapted to native patterns (e.g., *bljudce* 'saucer' borrowed with initial *#pl-*, reflecting the absence of voiced stops in Yupik), while loanwords from Russian in the Soviet era in the twentieth century are adopted without alteration (thus, *bl-* in 'saucer'), due to greater familiarity with Russian then, through exposure to Russian in the Soviet educational system. The same can be said for languages spoken throughout the Czarist Empire as opposed to the Soviet one.

Moreover, looking to the social setting of bilingualism helps to explain why in other Aromanian dialects, with presumably the same internal structural pressures of phonological oppositions as found in the above dialects, but spoken in a region with a different ambient other language, a different outcome occurred. In particular, northern Aromanian, with Macedonian as the dominant L2, has stops for the sounds in Greek loanwords that ended up with fricatives in southern Aromanian. For instance, Saramandu 1984: 432 and Cuvata 2006 give the following examples with *t* for Greek *θ*, *d* for Greek *ð*, and *g* for Greek *γ*, in (5.7), which contrast with the outcomes in (5.4):

(5.7) ana**t**emã (Grk ανάθεμα; see footnote 19)
dáscal^{u} 'teacher' (Grk δάσκαλος; Cuvata 2006: ***d****ascal*; *dispoti* (Grk δεσπότης))
grámă 'letter' (Grk γράμμα; Cuvata 2006: *agru* (Grk άγριος) *grambó* (Grk γαμβρός))

The effect in (5.7), we suggest, is due to the fact that in this region, the dominant second language that the Aromanian speakers know and are more familiar with is Macedonian, where the fricatives in question do not occur. Note also that this effect extends also to words with fricatives borrowed from Albanian, as in *dár**d**ă* 'pear,' from Albanian *dardhë*.

Of course, the proximate source of the loans is important here too; these borrowings could be through the mediation of the local variety of Macedonian, where Greek and Albanian fricatives would be reflected as stops, owing to Slavic phonological patterns (and relative unfamiliarity with Greek, we might add) at the time those words were borrowed from Greek into Slavic. This interpretation is bolstered by forms like *firidã* 'window,' ultimately from Grk θυρίδα. The inital *f* for the initial Greek voiceless fricative, found in some early loans from Greek into Slavic (note ChSl *anafema* in footnote 19, for instance, and *fimelj* 'foundation,' cited in Cuvata 2006, as opposed to Saramandu's form ***t****imél'*^{u}, from Grk θεμέλιο; see (5.4)), indicates the word might have come via a Slavic intermediary, although *firida* itself is not attested in Slavic. Moreover, the form ***t****imél'*^{u} 'foundation' corresponds to Mac *temel* 'idem.' Here the vocalization – the *i* in the first syllable in particular – points to northern Greek, although the /t/ in Saramandu's form (also Cuvata's /f/, see footnote 19) points to the possibility either of different Slavic intermediaries or northern

Aromanian native adaptations. The *d* in *firidã* could be a Slavic alteration of the Greek fricative but the word is absent from local Macedonian dialects, and so the adaptation could be native Aromanian, as in the 'pear' example cited above.

Another example for which dialectal data are relevant is δάσκαλος 'teacher, pedant,' which was borrowed as Alb *dhaskal*, now either archaic or pejorative, and Mac *daskal*, now either archaic or humorous. For Aromanian, the state borders correspond to the realizations of either /ð/ or not /ð/, i.e., Albania and Greece for the former, North Macedonia for the latter. In North Macedonia, however, while expected /d/ is the realization in most places, in Beala di Ghios (Mac Dolna Belica), a village where the inhabitants converted to Islam and have now mostly shifted to Albanian, the realization is *záscaľ*$^{\mu}$, which points to a native Aromanian tendency toward *z* in the absence of an *ð*-language like Greek as the dominant L2 and prior to the period when Albanian (with its *ð*) became dominant in the village.

Though moving the discussion away somewhat from Aromanian per se, one last consideration of realizations of Alb <dh> can be adduced. Mac *barz* 'goat or sheep with white spots on the head or body, person whose hair is going white' (Velkovska et al. 2003: s.v.), BSl dial. *bardza* 'fat woman,' *barzav* 'grey, ashen-colored' (BER I: s.v.), BCMS *barzast* 'partially black, white, or grey' (Skok 1971: s.v.), and Aro *bardzu, bardzã* 'black horse with brown hairs' (Cuvata 2006: s.v.), *bardzu* 'spotted' (BER I: s.v), 'white' (Skok 1971: s.v.), and Megl *bardză* 'white/piebald[?] goat' (Capidan 1935: s.v.), as well as Rmn *capră bardză* 'black piebald goat' (Skok 1971: s.v.), and *barză* 'stork' (Çabej 1976: s.v.), all go back to an etymon that is realized in modern Albanian as *bardh[ë]* 'white,' itself directly descended from IE **bhVrǵ*-, which has cognates in various other branches (Çabej 1976: s.v.). Given that later loans from Albanian into Aromanian show /d/, as they do in Slavic, the question here is whether the form of Albanian (or Proto- or Common Albanian) had [z] or [ð] at the time the word entered Slavic and Romance or whether Slavic and Romance speakers heard [ð] as [z] at that time. As Çabej (1976: s.v.) points out, the question remains open. We can note in passing that Macedonian has /dz/ for etymological /z/ under certain circumstances (e.g., *dzid* 'wall' from earlier *zid*), so that difference in the Aromanian and Macedonian forms could simply be a matter of contact dialectal realizations. The main conclusion from this example is that when discussing intimate borrowings between languages and dialects that have been in contact for almost two millennia, we must ask not only *whence* but also *when* (cf. footnote 21 and §5.4.4.3 for further discussion).

The crucial conclusion here is that there are Aromanian dialects in direct contact with Greek that have adopted Greek loans without phonological nativization, and this fact speaks to the relevance of the social surrounding, including the ambient second language, in which the borrowing occurs, in evaluating and accounting for phonological contact effects.[21] At the same time, it also speaks to time of contact.

21 Although we emphasize here the social setting, most of what we know about the circumstances under which Greek was learned by members of the Aromanian community during the Ottoman

Although these Aromanian examples should make the localization effects of bilingualism clear, we offer additional examples involving other languages in the Balkans in order to illustrate the pervasiveness of this phenomenon. Sandfeld 1930: 104 points out that Albanian dialects in Greece (i.e., Arvanitika) borrow Greek words with voiced fricatives as such, without altering the fricatives; for *ð* and *θ* and that this is not surprising since these sounds occur in native Albanian words. More interesting from the point of view taken here is the fact that this adoption strategy extends to the voiced velar fricative *γ*, not generally found elsewhere in Tosk Albanian, so that the Greek perfective stem αγαπησ- 'love' is borrowed into Arvanitika as *aγapis*.[22] In this context, the Aromanian of North Macedonia is interesting in that Greek *γ* is realized as [j] or [j̨] <y> only before /i/, reflecting Greek phonology, in some borrowed words (as is /g/ in some native words) in North Macedonia (Cuvata 2006: s.v.). Additional examples are the borrowing of Turkish high front rounded *ü* into Prizren Serbian (Remetić 1996) and the occurrence of Turkish mid front rounded *ö* in Turkish loans in some of the Romani dialects of eastern Bulgaria, e.g., Sevlievo and parts of Varna (RMS 2001–2005).

Bilingualism thus plays a crucial role in localized phonological diffusion. Importantly too, the effects can extend beyond loanword phonology. As Stankiewicz's remarks given above imply, phonological diffusion can follow intense lexical borrowing.[23] This is seen in cases where once-foreign sounds are extended outside of the loanword context in which they are adopted into the language (i.e., in unadapted form). For instance, Sandfeld 1930: 104 notes that southern Aromanian dialects have [*γ*] for *g* in words of Slavic origin, e.g., *aγunesku* 'chase' ultimately from Slavic *goniti* (cf. Rmn *gonesc*, with [*g*]) and Capidan 1940 gives cases of Latinate words in some Aromanian dialects that take on the Greek fricatives, e.g., *ðimtu* 'wind' for the more usual and widespread *vimtu*, from Latin *ventus*.[24] Moreover, Latin *femina* 'feminine, female,' gives *feamín* in Macedonia (Cuvata 2006: s.v.) but *θeamin* in Greece (Papahagi 1974: s.v.) and Albania (Neiescu 1997: 73). Similarly, as noted above, some Serbian dialects in Kosovo, such as that of Prizren, frequently have *ü* as a result of contact with Albanian and Turkish (Remetić 1996: 356, 366–367). Although /ü/ occurs exclusively in loanwords, Remetić 1996: 366 argues that it is part of the phonemic system because it

period, and before, is inferential. That is, the observation in Récatas 1934 (see §3.0 and footnote 2 therein) that Aromanian women knew little Greek suggests that males in Aromanian villages in Greece were not exposed to much Greek before beginning to move around outside of the home setting. Still, the men could have gained a command of Greek at a relatively early age and that could explain their phonological "success" (i.e., nonnativization) with Greek borrowings. Nonetheless, this early-exposure explanation is itself inferential, based on the assumption, consistent with the critical period hypothesis (see §3.2.1.5, footnote 42), that only early exposure to a language can ensure the occurrence of unassimilated loanwords in a borrowing language. See footnote 25 below for a different, more socially based, interpretation.

22 Elsewhere in Albanian, the verb *agjapis* 'love' occurs, with a nativized *gj* for the Greek *γ*; Newmark 1998: s.v., labels it as a nonstandard form.

23 And recall here Thomason & Kaufman 1988 and their scale of borrowing (cf. above §5.2).

24 A Greek-like interdental fricative replacing here a labiodental fricative is understandable in terms of the acoustic similarity of the sounds in question.

can be replaced by /u/ or /i/, but /u/ and /i/ in native words are never replaced by /ü/. We can add that there are also some near-minimal pairs, e.g., *düšema* 'rug'/*dušama* 'soul (DAT. PL.)', *Ümere* 'Ümer (man's name, VOC.)'/*umeri* 'moderates, limits (3SG). Similarly, in the Macedonian dialect of Boboshtica (Mac Boboščica), /ü/ occurs in words of Turkish and Albanian origin (the former may have entered through an Albanian intermediary), but it has also been extended to some native Slavic words, e.g., *klüč* 'key' (Vidoeski 2000: 247). The same is true for some Aromanian dialects in Albania (Neiescu 1997: 102).

In such cases, familiarity with the other language makes these originally foreign sounds less foreign-seeming, as indicated by Remetić 1996: 366 in his description of Prizren multilingualism.[25] We can view the extension of these sounds into native words or loanwords of a different origin as a sort of hypercorrection or hyperaccommodation. Keeping in mind that there would be different degrees of familiarity with a second language among speakers in a given speech community, if there were some speakers who fluctuated between native (or nativized) *g*, for instance, and foreign *γ* in their pronunciation of a loanword where the source language had *γ*, or even if there were some speakers who had only *g* and some who had only *γ* in a given word, then the alternation between *g* and *γ* in individual speakers or across sets of speakers could be the basis for the extension of the novel sound into words which originally (from an etymological standpoint) had *g*. Thus bilingualism again is a contributory force, and one can imagine that such a scenario would be possible only in an area where there were bilinguals (of differing abilities) and where the first appearance of the innovative foreign sounds is in unadapted loanwords.[26]

Finally, to close this section of localized contact-phonology case-studies, and by way again of emphasizing the social dimension to the borrowing of phonology, we examine a case where there is no phonological convergence and consider causes for stability of certain sounds even under intense contact. In particular, in Balkan Romani, as Friedman 2001a points out, the Romani voiceless aspirated consonants (whether from earlier voiceless or voiced aspirates, or from other developments, e.g., involving clusters) are maintained intact and remain distinctive, even though

25 Thus, there is a possible ideological dimension here, in that speakers' perceptions of foreignness and the degree of comfort they feel with particular foreign elements, and how they act on these perceptions, may well matter for the outcome. Neikirk-Schuler 1996 offers a highly relevant approach, discussing morphological nativization in Balkan Slavic as an act of "staking out linguistic territory" as speakers assess loanwords through the filter of a featural assignment of [±MINE] or conversely, and more tellingly, [±NOT MINE]. Similarly then, the act of bringing a foreign pronunciation into one's linguistic system can be an ideological act, with speakers in effect saying that a particular sound is not too foreign to adopt, not so external to the system that it cannot be considered [-NOT MINE] (= [+MINE]). The fact of originally borrowed phones entering native vocabulary suggests there may be something more to the introduction of these sounds than just speakers' ability or willingness to sound foreign in their pronunciation. See footnote 21 above, however, for a different perspective.

26 From observable contemporary situations, it is known that speakers can use unadapted loanwords to generate particular effects (such as sounding learnèd or cosmopolitan) and in general to contribute to the construction of a particular identity or persona. What we do not know, and probably can never know, is the effect speakers in these pre-modern Balkan contexts had, or were aiming at, through their use of unadapted loans.

all surrounding languages in the Balkans lack distinctive aspirates.[27] Romani has been in sustained and intense contact with all the languages of the Balkans, and Romani speakers are bi- or multilingual in their native Romani and other co-territorial languages. One might well suppose that these sounds, unusual from a pan-Balkan perspective, would be particularly vulnerable in Romani, but such is not the case; there is no phonological effect on Romani from bilingualism here, as indicated by the representative forms in (5.8), some forming minimal pairs with words with unaspirated stops:

(5.8) khel 'play' (earlier Indic *kh-* from *kr-*, cf. Skt *krīḍ-*)
kher 'house' (vs. *ker* 'do!')
čhorel 'spill, empty' (vs. *čorel* 'steal')
the 'and'(vs. *te* 'DMS')
phral 'brother' (earlier Indic *bh-*, cf. Skt *bhrātar-*)
phal 'pale' (earlier Indic *ph-*, cf. Skt *phalaka-* 'fruitful')
pherel 'fill' (vs. *perel* 'fall')

At the same time, however, since aspiration is limited to native vocabulary items, the feature sometimes serves to distinguish nonnative from native items: *čaj* 'tea' / *čhaj* 'girl', *kula* 'tower' / *khula* 'nonsense.'

But here the social isolation of Romani speakers, being as they are on the margins of Balkan society, is the crucial factor; even though Roms are bi- and multilingual as a rule, Romani bilingualism is unidirectional – non-Roms do not in general learn Romani whereas Roms have to learn other languages (cf. Friedman 2000d). Keeping certain aspects of Romani phonology distinct from the other languages and maintaining a phonological boundary between it and neighboring languages, then, becomes a way for Romani speakers to maintain a social distance between themselves and other speakers and to demarcate Romani through a distinctive feature in its phonology. In a sense then, the phonology in this case iconically marks this distance – these sounds, marked and marginal within the overall context of Balkan phonology, are emblematic of the marked and marginal status of Romani speakers from a societal perspective.[28]

27 Ancient Greek had voiceless aspirates, but they developed into fricatives in the Hellenistic period long before Romani entered the Balkans. There is one exception among modern languages: Tsakonian Greek (see Pernot 1934) has distinctive voiceless aspirates; these do not continue the Ancient Greek sounds but rather are conditioned outcomes of other earlier sounds. The relative geographic isolation of Tsakonian in Greece into the early twentieth century parallels the social isolation of Romani speakers. Furthermore, there are other ways in which aspiration is not totally alien to Balkan sound systems: Turkish shows allophonic aspiration, with voiceless stops generally aspirated word-initially but not elsewhere, and detailed instrumental phonetic investigation of Greek via spectrograms reveals occasional, but inconsistent, light aspiration on word-initial voiceless stops.

28 Indeed, similar facts concerning the maintenance of voiceless aspirates are found throughout Romani, even outside of the Balkans. Rather than vitiating this example, this observation demonstrates that the social context is key, since the Balkan Romani social facts of marginalization and unidirectional bi- and multilingualism are replicated across the Romani-speaking world.

5.4 Proposed Phonological Balkanisms: A Survey

With a specific mechanism – bi- and multilingualism – and the overriding importance of social setting now established as a basis for understanding contact-induced phonological effects, we can better consider the specific features proposed in the literature as phonological Balkanisms, organizing the examination of them by phonological parameters (vowels, consonants, subclasses of consonants, etc.) and distribution (possibly pan-Balkan versus restricted to a relatively small geographically contiguous region, or a small number of languages). As the ensuing discussion reveals, all of the features we consider must also be understood in the context of what is known about phonetic naturalness[29] and about the historical phonology of the particular languages, in addition to, as always in the Balkans, pervasive, generally mutual, bi- and multilingualism and the use of linguistic means in assertions of various group identities.

It is important with regard to the phonological features, since their very status as Balkanisms is controversial in ways that the status of commonly agreed upon morphosyntactic or lexical convergences is not, that their presentation be accompanied by some indication of who has adduced them. Thus, even though these features are generally mentioned in a number of handbooks, of which a selection is utilized here, we nonetheless indicate their source, identifying as well, where possible, the originator of the observation (adding other references beyond these as needed); we use the following abbreviations, placed in square brackets after the feature:

FM = Miklosich 1862
KS = Sandfeld 1930 (pp. 102–104, 114–115, 124–127, 145–146)
EA = Afendras 1968
HS = Schaller 1975 (pp. 124–133)
EB = Banfi 1985 (pp. 45–51)
JF = Feuillet 1986 (pp. 45–53)
PA = Asenova 2002 (pp. 28–42)
SD = Sh. Demiraj 2004 (pp. 83–94)
IS = Sawicka 1997
SV = Steinke & Vraciu 1999 (pp. 97–105)
UH = Hinrichs 1999b (pp. 435–437)

Some of these proposed phonological Balkanisms are diachronic in nature, reflecting sound changes that various of the languages have in common or that at least yielded parallel results. Others are simply (static) synchronic facts, distributional or otherwise, about the language(s) in question, which cannot be taken as Balkanisms unless some convergent mechanism is identified. In either case, we offer here

29 We refer to "naturalness" fully recognizing that it is a difficult notion to define. We rely on a commonsense view of naturalness here – for instance, assimilations involving features in adjacent segments certainly qualify – though where possible, we point to parallels in noncontact situations as a control over the effects of phonetic and phonological naturalness vs. contact-induced causation.

sufficient information to evaluate claims made about the features. For the most part, we believe that language contact is involved in most of these convergent features, though in some instances, the evidence is only circumstantial in nature, based on geography rather than on the occurrence of some detail in all the languages involved. We try to give a frank appraisal of the viability of each feature as a Balkanism, whether of wide or localistic distribution.

5.4.1 (Nearly) Pan-Balkan Features: Vowels

With respect to vowels, the features in (5.9a–f) figure prominently in discussions of Balkan sprachbund phonology:

(5.9) Convergences involving vowels

a. A five-vowel (sub)system of *i – e – a – o – u* in each language [EA, HS, EB, JF, PA, IS, UH] (drawing on Havránek 1933)
b. No length opposition in vowels [EA, HS, EB, JF, PA, SD, IS, UH] (drawing on Havránek 1933)
c. No rounded/unrounded opposition in high front vowels [EA, HS, JF, PA, IS]
d. No oral/nasal opposition in vowels [EA, HS, EB, JF, PA, IS, UH]
e. No diphthongs (except in Romanian) [JF]
f. Existence of an (accented) (mid-)central (unrounded), i.e., schwa-like, vowel [FM, HS, EB, JF, PA, SD, IS, UH]
g. Vowel reduction in unaccented syllables, with consequent alternations between accented and unaccented forms [FM, KS, EA, EB, JF, PA, IS, UH]

Feature (5.9g) is often discussed in such a way as to suggest a widespread Balkan character to it (e.g., for UH, it merits a whole paragraph in his two-page discussion whereas other more widespread traits are given just a line); still, we defer full treatment of it to the later section on more regionally restricted features (for reasons that become apparent from our discussion).

The first four of these, (5.9a–d), are widely agreed upon and refer to the fact that most of the languages show a five-vowel "triangular" system either as the whole of the vowel space, as in most of Greek (including the standard language), or as a significant subsystem within the overall system, as in Balkan Slavic, Tosk Albanian, Romanian, and others. Moreover, these traits mean that these vowels are relatively "clear," to use Sawicka's term, without any of what might be termed "overlay" features such as length or nasality or, for the front vowels, rounding, and for the mid-vowels relative height (open/closed) that secondarily differentiate basic vowel qualities. Feuillet's further suggestion, unique to him it seems, of no diphthongs, (5.9e), can be seen as in keeping with a sense of a relatively simple and unadorned system of vocalic elements. The schwa observation dates to FM, and is generally stated with reference to a stressed vowel. Moreover, it is incompatible with the five-vowel notion, since it is always a sixth vowel.

As with most putative Balkanisms, these features, to the extent they are even valid, represent movement away from earlier systems that were quite different. For instance, counter to (5.9abce), Ancient Greek had both *i* and *ü* and thus a rounding

distinction in high front vowels; short and long vowels, e.g., *i* and *ī*, a length distinction; in some dialects and at some stages of development more height distinctions than in Modern Greek, e.g., *ī ē ǣ* and several diphthongs, e.g., *üj oj aw*; also, earlier Slavic, as represented by Old Church Slavonic, had an oral/nasal distinction, counter to (5.9d), contrasting oral *e o* with nasal *ę ǫ*. Both Ancient Greek and Common Slavic had tonal systems, Common Slavic also had distinctive length and diphthongs (although these were lost at the end of the Common Slavic period and then reintroduced), and Latin too had distinctive length, a feature that is likewise reconstructed for Common Albanian. There is also evidence that Old Church Slavonic (OCS) may have had a high front rounded vowel; similarly, Common Albanian can be assumed to have had *ü*, as this vowel is found in all Albanian dialects except southernmost Tosk in Albania, and both Arvanitika and Arbëresh (and in all of these, contact with Greek or Italian may have played a role in the absence of *ü*).

As it turns out, though, there are serious problems associated with each of these features, thus rendering their value as true Balkanisms quite dubious. For one thing, they are not found in all of the languages. Feuillet does point this out (pp. 45–47), and it could be argued that in this regard these features are no different from the morphosyntactic and syntactic features widely recognized as Balkanisms but also not realized uniformly in all the languages (see Chapters 6 and 7). Nonetheless, this fact alone might be thought to diminish the value of these features somewhat as indicative (or demarcative) of a sprachbund, at least under the usual conception of the construct as defined by a set of widely distributed shared features. On the other hand, a distribution of features like this would be entirely in keeping with the localistic, cluster approach to the Balkan sprachbund advocated in §3.3 and in §5.2 and §5.3.

However, there are other problems as well, and these concern issues of phonetic and typological naturalness, historical interpretation of the facts, and certain definitional matters. In what follows we clarify the language and dialect distribution of these features and discuss the many other problematic aspects connected to them. Some of these observations have been made by others, and sometimes doubts about the value of these features as Balkanisms were raised almost from the start. Sandfeld 1930: 125, for instance, mentions the Albanian/Romanian supposedly shared development of *a* to *ə* (relevant to (5.9f)), but then immediately dismisses it, noting that "*le passage de a en ă est bien plus régulier en roumain . . . [mais] ë en albanais représente aussi d'autres voyelles atones que a*" ('the passage of *a* to *ă* is indeed more regular in Romanian . . . [but] *ë* in Albanian represents as well unstressed vowels other than *a*'). Such concerns may well contribute to the seeming sense of discomfort (see §5.1) that has pervaded many previous discussions about phonological Balkanisms, and thus they deserve enumeration and elaboration.

5.4.1.1 Distribution of Balkan Vowel Features

With regard to the distribution of these features around the Balkans, certainly the most comprehensive survey of the details of Balkan vowel systems is to be found in the excellent work of Sawicka 1997: 13–30. A key point to be extracted from her discussion is that the dialects show distinctions that are not found at all in the standard languages; thus, basing assessments of Balkan phonology on standard languages, as most accounts do, gives a very misleading picture. Similarly, attention to phonetic detail can alter the picture somewhat. Summarizing in large part from Sawicka (though Friedman 2006a is the source in (v)), when one takes a fuller range of material into consideration, several essential clarifications emerge:

i. regarding (5.9a), more elaborate vowel height oppositions occur in various Balkan dialects. For instance, in the Greek of Thessaly, as reported by Tzartzanos 1909, a low front *æ* occurs, along with high *i* and mid *e*. Distinctive low front *æ* occurs in a number of peripheral Macedonian dialects as well as in Teteven-Erkeč in eastern Bulgaria. Distinctions of open and closed mid-vowels are also found in various Macedonian and Bulgarian dialects as well as Meglenoromanian (cf. also Gjinari 2007: 86 on Albanian and Koneski & Vidoeski 1983: 111–134 for Macedonian).

ii. regarding (5.9b), some languages in the Balkans do distinguish long and short vowels; while it is true that the absence of length distinctions is one of the defining marks of the Torlak dialects of BCMS, vowel length (Gjinari 1989: 111) is preserved in all of Geg Albanian and in southern Tosk (Çam and Lab, except around Përmet), and now occurs secondarily even in other Tosk dialects under conditions for compensatory lengthening (e.g., some word-internal vowels become distinctively long with the loss of final *-ë*, as in *mīr* from *mirë* ‘good,’ even in the standard language (Sawicka 1997: 24)), and via contractions in Balkan Turkish (e.g., *alām* from *alayım* ‘that I get’) and Macedonian (even colloquially in the standard language, e.g., *sāt* ‘clock’ from *saat*, *padnā* ‘they fall’ from *padna-a*, and dialectally, e.g., in the Dojran dialect *vō* ‘water’ from *voa*, itself from *voda* through the occasional lenition and loss of intervocalic *d* (Sawicka 1997: 25, 29)).[30] We note as well that Greek vowels, even in the standard language, can be protracted as a hesitation noise, as is commonly heard in the discourse marker δηλαδή, pronounced as [ðilaðī], ‘that is to say’ (where the final vowel can be “overlong,” i.e., protracted for several morae, thus perhaps better represented as [ðilaðiii . . .]).

The secondary and thus more recent nature, or, in the case of Greek hesitation lengthening, the iconicity and marginality, of these long vowels could of course signal that claims such as (5.9b) are valid for a pre-contemporary stage of the Balkan languages. However, formulations of Balkanisms in the literature

30 Sawicka 1997: 27, however, is quick to point out that these realizations are variable and “the contemporary situation does not allow us to speak about the long vocalic phonemes in the Standard Macedonian system, but [a] tendency towards their phonologization is quite obvious.”

usually are not temporally restricted, and generally refer to the facts from contemporary standard languages. Thus it becomes important to judge these claims against the fullest range of data available to see just how valid they really are.

iii. regarding (5.9c), a rounded/unrounded opposition in high front vowels occurs in most Albanian dialects and also in Balkan Turkish (Sawicka 1997: 13)[31] and in southern Montenegrin dialects. The fuller story here is important, though. The relevant Montenegrin dialects are in close contact with Albanian, where /y/ is found, but both East Central Geg (found mostly in North Macedonia – and, moreover, in contact with Macedonian villages even on the Albanian side of the political border) and southernmost Tosk (especially the Lab, Çam, and Arvanitika dialects) show an unrounding of *ü* to *i*, presumably due to contact with Macedonian, Aromanian, and/or Greek, depending on the particular region. Greek also merged *ü* with *i*, though this change was completed only around the tenth century (Newton 1972: 19) and maybe even as late as the nineteenth century to judge from some descriptions of Greek of that era (Pantelidis 2020). Similarly, if OCS developed *ü, ǒ* (from *j* + *u, ǫ*), a matter of some controversy (see Lunt 2001: 25–26, 30), it must have quickly merged with *u*, *ǫ*, as no later South Slavic language attests this vowel. Still, as Sawicka 1997: 16 observes, there are dialects of Greek (e.g., parts of Thessaly, Greek Macedonia, and Thrace, see Newton 1972: 46–52 for details) where front rounded *ö* and *ü* develop secondarily from sequences of *io* and *iu*; moreover, in some of these dialects, Turkish loans with front rounded vowels are adopted without alteration – Newton (p. 50) for instance cites μπαλντ[ü]ρς ([baldyrs]) 'vagabond' as a Turkism in the Tyrnavo dialect of Thessaly.[32] And, Ronzevalle 1911, 1912 makes it clear that the Greek of Ottoman-era Edirne borrowed many Turkish words with *ü*, such as *džüdžes* 'dwarf' (Turkish *cüce*). Moreover,

31 In West Rumelian Turkish, however, word-final *ü* is unrounded to *i*. Note further that some changes that look "Balkanesque" in Balkan Turkish are not to be interpreted as such: while the elimination of *ö* in some dialects might seem like the loss of a labiality opposition in front vowels, Sawicka 1997: 13 states that "in most instances it changed into ü, rarely into o," so that distinctive labiality was maintained.

32 Newton gives only this word as an example of [ü], though Tzartzanos 1909, Newton's source, gives other examples. Still, Newton's lone example is a somewhat curious one, as it is not cited in the large Modern Greek dictionaries now available (Babiniotis 1998, LKN, Charalambakis 2014). Its absence from contemporary sources may be due to its being an "occasionalism" in a regional dialect from some 100 years ago. On the Turkish side, the standard form is (now) *baldır* (Redhouse 1968) but there is a dialectal (and older) form *baldur*, that could well have been the input into the Tyrnavo form. The final -ρς in the Greek may reflect the Turkish devoiced word-final /r/, or simply the addition of a Greek (masculine) nominative ending. The semantics of *baldır/baldur* do not match the Greek, meaning 'calf (of the leg); stem (of a plant)' but in one phrase, *baldırı çıplak*, meaning literally 'calf-naked,' and thus 'barelegged,' and by extension 'a rowdy, ruffian,' i.e., a 'ragamuffin,' it approaches the Greek, if extracted from the phrase. The front rounded *ü* is unexpected, given the Turkish forms, but if not just an unrecorded Turkish dialect feature, it could be a hyperforeignism (see Janda et al. 1994), as suggested in Joseph 2011c, based on the occurrence of *ü* in other Turkish words (assuming some degree of familiarity with Turkish in general in Thessaly at that time; see also §5.3).

some Aromanian dialects in Albania also show *ü*, at least in loanwords (Neiescu 1997: 102, 314).

iv. regarding (5.9d), an oral/nasal opposition in vowels occurs in all dialects of Geg Albanian except the town of Debar (Dibra), and remnants of nasality are to be found in extreme southwestern Macedonian (the Kostur (Grk Kastoria) and Korça regions) and in Suho-Visoka (Grk Sokhos and Óssa) Macedonian (in the southeast, near Thessaloniki). In Macedonian these remnants are not nasal vowels per se but rather nasal consonants that represent "segmentalization" (in the terminology of Hock 1991: 120) of the nasality feature of earlier nasal vowels; still, even though the loss of nasality in much of the Macedonian-speaking territory occurred as early as the thirteenth century (Koneski & Vidoevski 1983: 36), such developments in outlying areas have been taken to suggest a persistence of nasality until relatively recently in some of Macedonian, but on the other hand, they could represent early contact-induced denasalization (see Friedman 2018a and references cited therein). Moreover, nasal or semi-nasalized vowels occur in some parts of the Balkans innovatively, for instance, in eastern Montenegrin dialects in contact with Albanian (Stevanović 1935: 17–18). Further, though not (necessarily) distinctive, vowel nasalization is reported for some speakers for the realization of vowels with nasals in a few modern Balkan languages. In the Judezmo of Thessaloniki, for instance, according to Afendras 1968: 96 (drawing on Crews 1935), sequences of *a*/*o* + *n* develop into "nasalized vowels in final position." In Greek, according to Arvaniti & Joseph 2000, this development is found before stops, e.g., [ãbeli] for αμπέλι 'vine,' and in Bulgarian, Scatton 1993: 190 reports that "vowels are nasalized before nasal consonants followed by fricatives." Interestingly, too, Scatton adds that "often the nasal consonant is lost," a further development which would create some distinctive and contrastive nasalized vowels; he cites the demonstrative /onzi/ 'that.M.SG' with the realizations [õnzi] and [õzi]. Still, in the Greek case, the nasal may be considered to be underlyingly present at least as far as the Greek diasystem is concerned, and in the Bulgarian case, there is support for recognizing an underlying consonantal /n/ in /onzi/ based on related forms of the demonstrative. Thus the nasalized vowels here may be phonetic only and not (fully) phonologized. Moreover, these developments are presumably temporally quite removed from any (significant) Balkanological contact and period of convergence, and as such they may be of little consequence for evaluating claims about historical Balkanisms; they show that judging the validity of Balkanisms solely on the basis of evidence from contemporary standard language data is potentially problematic.

v. regarding (5.9f), a stressed mid-to-high central vowel is certainly present in several languages, most notably Tosk Albanian (signaled with the *ë* grapheme), Torlak BCMS, and most of Bulgarian (signaled with the <ъ> grapheme), and most Macedonian dialects (where it is written with an apostrophe in Cyrillic and *ă* in transliterations), as well as Romanian, where there is both

a high and mid central vowel (signaled with the *î* and *ă* graphemes respectively) and Aromanian, where the southern dialects have the high versus mid central vowel opposition and the northern dialects (standard Aromanian of North Macedonia <ã>) as is also the case in Meglenoromanian, have just the mid central vowel. It also occurs in some Romani and Judezmo dialects. Thus schwa is not pan-Balkan but its distribution does cover broad areas of the Balkans. Moreover, it is found in some languages/dialects not usually thought of or mentioned in this context; some scholars (e.g., Naylor 1980) have argued that on structural grounds there are more sounds, including the syllabic *ŗ* of Macedonian and BCMS and the back unrounded *ɪ* of Turkish, that should be considered as relevant to this putative Balkanism. The relevant facts concerning the distribution of this feature outside of Tosk, Bulgarian, and Balkan Romance are presented well in Friedman 2006a so we quote from it here *in extenso*:

> A phenomenon common in the Balkans is the existence of a stressed schwa, but its status as a contact-induced phenomenon is not pan-Balkan. Greek lacks stressed schwa altogether. In Macedonian, almost all the dialects outside the west-central area have stressed schwa, but of different origins in different areas, and some western peripheral dialects also lack stressed schwa. Most of Bulgarian has stressed schwa, but not the Teteven-Erkeč and central Rhodopian dialects. In Albanian, stressed schwa develops from nasal *â* only in Tosk, but it is incorrect to characterize all of Gheg as lacking stressed schwa, since it also occurs in central Gheg as a result of later processes of diphthongization. Thus, for example, in Mirdita stressed /i/ and /î/ are diphthongized (and denasalized in the case of /î/) to /ej/, then centralized to /əj/ which can be monophthongized to /ə/ in words such as *korrëk* (Standard *korrik*) 'July,' *mullë* 'mill' (Standard Gheg *mullî*). Moreover, Turkish /ɪ/ can be realized as schwa in loans, e.g., *açëk* < Turk[ish] *açık* 'open,' a phenomenon occurring in many Balkan dialects. Romani has schwa when in contact with languages that have it, and some native words in some dialects also develop stressed schwa, e.g., *sastrən* 'iron' from *sastrn* via influence of Bulgarian or Romanian phonotactics. In some WRT, there is a tendency to lower and front the high back unrounded vowel bringing it closer or all the way to schwa.

Even with these facts, some of which are more appropriately considered as local features found on a localized basis, it becomes apparent below (in §5.4.1.6, and see also §5.4.4.3 and §5.4.5.1) that there is reason to be skeptical about this feature as a pan-Balkan one. Judezmo is unique in this respect, as those dialects with stressed schwa occur exclusively in Bulgaria, and it is clearly contact-based rather than an independent development (Quintana Rodríguez 2006: 359).

Taking some of these points together, it turns out therefore that there are Balkan dialects with extraordinarily rich vowel systems in contrast to statements about simple clear five-vowel systems. As Sawicka 1997: 14 points out, Beci 1981, 1984 observes that some Albanian dialects, e.g., that of Shkodra and environs, have as many as twenty-four vocalic phonemes, with length and nasalization distinctions, e.g., *i – iː – ĩ – ĩː*.

5.4.1.2 Definitional Issues

Issues of a definitional nature plague (5.9e), the claim regarding diphthongs in the Balkans, and these go beyond the mere need for an accurate statement of the feature's distribution. Diphthongs are generally taken to be sequences of a vowel plus a semi-vowel (glide) with a single syllabic peak, e.g., [ej], [aw], and the like; there can be ambiguity, however, in regard to the level of analysis at which to apply such a definition. That is, one could refer to only surface realizations, to underlying phonemic representations, or even to morphophonemic representations, in declaring a given vocalic grouping to be diphthongal or not. This analytic ambiguity can be illustrated even for Romanian, the language that Feuillet himself admits is outside the scope of his claim for the Balkans.

In particular, a noncontrastive surface diphthong, such as the [aj] in Romanian *mai* 'more' or the [uj] of *cuib* 'nest,' could be taken at face value and treated as underlying diphthongs, thus /maj/, /kujb/. Or, since there is no contrast with a (presumably disyllabic) two-vowel pronunciation [ma(.)i]/[ku(.)ib], they could be analyzed respectively as /mai/ (or /ma.i/) and /kuib/ (or /ku.ib/) underlyingly, with a phonological rule of glide-formation being responsible for the surface diphthong. And, some surface diphthongs in Romanian, e.g., [ow] in *birou* 'office, desk,' alternate with superficial two-vowel sequences of [o.u] in different syllables, as in the plural *birouri* ([bi.ro.ur^{j}]), so that the singular can be taken on the basis of its morphophonemic behavior to be underlyingly nondiphthongal, / . . . o.u . . . /. And even orthographic <ea oa>, the most commonly cited Romanian diphthongs in the Balkanological literature (and quite possibly what Feuillet had in mind when excluding Romanian from the scope of [5.9e]), could conceivably be treated as underlying vowel–vowel sequences, and thus diphthongs only superficially, not underlyingly.

Still, even if it might be possible to bring Romanian in under the rubric of (5.9e) through some such analytic devices and juggling of levels of analysis involved,[33] and possibly so also for other Balkan languages, in many instances, in languages for which Feuillet did intend (5.9e) to be applicable, the analytic sleights-of-hand fail. Thus, in (standard) Albanian there is a contrast, for at least some speakers and at least in careful speech, one reflected in the standard orthography,[34] between the vowel sequence [oi] and the diphthong [oj] in final position, as in the 3SG past form

33 We do not pretend to offer here a full treatment of the phonology of diphthongs in Romanian; these remarks are merely suggestive of how to proceed if one were so inclined. Petrucci 1999 is a treatment of Romanian phonology that includes discussion of such analytic issues, and, interestingly, considers the language to have distinctive diphthongs (see especially his p. 42). Similarly, Chitoran 2002a is a more theoretically oriented work but covers the phonology of Romanian quite thoroughly, ultimately arguing for underlying diphthongal representations for <ea>, <oa>, though not for <ai>, <ui>, etc.

34 In fast and/or casual speech-styles, this distinction generally is neutralized in favor of a diphthongal pronunciation, offering another reason for caution in applying (5.9e) – at best it might apply only to certain speech registers or tempos for some standard languages but cannot be taken as an accurate reflection of pan-Balkan phonetics.

këndoi ‘s/he sang’ versus the 1SG present form *këndoj* ‘I sing’; Sawicka 1997: 27 discusses a similar case in Macedonian (e.g., regarding [moi] ‘my.PL’ versus [moj] ‘my.SG’), and one can note contrasts in Bulgarian such as singular [muzej] ‘museum’ versus plural [muzei].

In some languages, moreover, accentual behavior points to true diphthongal status for at least some vowel + glide sequences, making it impossible to resolve them underlyingly into vowel + vowel sequences. Thus in Macedonian dialects, such as that of Tetovo, that have a strict and inviolable antepenultimate stress placement rule, one finds diphthongs occupying the penultimate position, as in the verbal adverb *čítajkji*, with [aj] ‘(while) reading.’[35] And, in Greek, the otherwise quite strictly observed generalization that no accent can occur farther from the end of the word than the antepenult (for the standard language at least)[36] would be jeopardized by the stress placement in words like γάιδαρος ([γájðaros]) ‘donkey’ or χάιδεμα [xájðema]) ‘caress,’ if they were taken to have /a.i/ sequences underlyingly. Further, the *accentual adjustment*[37] that one finds with the addition of postpositive possessive pronouns in Greek, e.g., όνομα ‘name’ ~ όνομά μου ‘my name’ (see (5.14)/(5.15) below in §5.5.2), demonstrates that some penultimate diphthongs cannot be easily resolved into vowel + vowel sequences; that is, κορόιδο ([korójðo]) ‘laughing-stock’ with a possessive is accentually unchanged, κορόιδο μου ([korójðo mu]) ‘my laughing-stock’, whereas if [oj] were disyllabic underlyingly (so that κορόιδο was /koroiðo/, one would expect to find (counterfactually) κορὸιδό μου* ([korὸjðó mu*]) or κορὸιδό μου* ([korὸiðó mu*]).[38]

35 The corresponding form in the standard language is *čitájkji*, a surface exception to the otherwise general antepenultimate accent placement. For the standard language, this form could be analyzed as underlyingly /či.ta.i.ḱi/ (and historically it is *čitáekji*) but such an analysis for Tetovo Macedonian is possible only at the expense of an otherwise regular stress placement generalization.

36 Some regional dialects allow accent on the fourth or even fifth syllable from word-end. The status of potential diphthongs in such dialects remains an investigation for the future.

37 This is our term, and we use it here and in §5.5.2 as a way of generalizing, for presentational purposes, over the Greek developments given here and developments that are called *double accent* (cf. Alexander 2004) in some Balkan Slavic varieties. We do recognize important differences between Greek and Balkan Slavic in this regard but generalize, nonetheless, since in both Greek and Balkan Slavic there are changes – i.e., adjustments – in the accentual patterns associated with extended prosodic domains.

38 We say “easily” because it is always possible with devices such as rule ordering (or some mechanism with a similar effect, e.g., constraint rankings within Optimality Theory) to avoid problems noted here with particular lexical items. For instance, stress placement in Greek could simply be stipulated to follow diphthongization in deriving χάιδεμα ([xájðema]). Such an account, though, has to be lexically specific: a verb like μπαΐλντίζω ‘faint’ (from the Turkish root *bayıl-*) has a surface diphthong in the present tense form given here; however if that diphthong is from an underlying vowel sequence (thus, /baildizo/), then in the case of the past tense, with its characteristic antepenultimate stress, stress placement would have to precede diphthongization, since the form is μπαΐλντιζα ([baíldiza]) ‘I was fainting,’ not μπάΐλντιζα* ([bájldiza*]) (see Mackridge 1985: 31 and Baltazani & Topintzi 2012 for similar argumentation). We note that Macedonian has surface sequences that seem to be diphthongs phonetically, e.g., [pawza] ‘pause, break’ but the antepenultimate stress placement rule in that language shows such forms to have underlying heterosyllabic vowel sequences (/pa.u.za/); the definite form of ‘pause’ is [pa.ú.za.ta] and not the [páw.za.ta] that an underlying diphthong analysis predicts. Cf. also *váučer/vaúčeri* ‘voucher SG/PL.’

It can be concluded therefore that (5.9e) simply is not an accurate statement regarding Balkan phonology, neither for the standard languages taken together nor for various dialects.[39]

5.4.1.3 Balkan Vowel Features and Naturalness

Several of the features cited above fall short of being compelling as potential Balkanisms by not measuring up against a criterion of naturalness.[40] That is, if from a cross-linguistic standpoint, they are quite common or occur frequently or if they represent phonetically trivial phenomena, then their presence in any language – in the Balkans or elsewhere – is not particularly noteworthy but rather is just attributable to the nature of language in general. In such instances, there would be reason for doubting their probative value for the sprachbund, as they could easily have arisen independently in any given language, without contact with speakers of other languages being a contributing factor.

Among the vowel features of this sort are (5.9a), since a five-vowel (sub)system of *i – e – a – o – u* is an exceedingly common array of vowels, as noted by Schwartz et al. 1997ab, who observe, based on the 317 languages in the UCLA Phonological Segmental Inventory Database (UPSID), that it is found in 97 percent of the 100 languages with a five-vowel system; (5.9f), since schwa is a not-infrequent sound that occurs in many non-Balkan languages (e.g., Malay, English, Biblical Hebrew, Kabardian, and Komi, to name a few), though admittedly if the feature is further specified as referring only to accented schwa, it might be less common; and (5.9g), since some sort of phonetic reduction of vowels (e.g., to a less distinct vowel such as a schwa or to a vowel reduced in duration such as a high vowel, which has an inherent shorter duration than a lower vowel), or phonological reduction (e.g., to one of a smaller set of phonological contrasts) in unaccented syllables is not at all unusual cross-linguistically, although the particular form that such reduction takes in various Balkan languages may well be more noteworthy (see §5.4.3.9).

This reasoning can also be applied to sounds that were eliminated from Balkan languages. Such sounds might be suspected of having suffered their fate due to language contact, yet in some cases they turn out to be rare or at least infrequent cross-linguistically. As such, they can be considered marked, inviting the inference that their loss in a contact situation is not at all unexpected. One such sound is the high front rounded vowel *ü* (IPA [y]), which is not widely found in languages of the world (front rounded vowels occur in only about 9 percent of the world's languages surveyed in Maddieson 1984: 248–251, for instance), and vowel nasalization could be another such case. Whether such a principle of simplification as an outcome in

39 Sawicka 1997: 26–30 cites numerous cases of secondarily created diphthongs, e.g., through the loss of intervocalic consonants, in various Macedonian dialects, and more generally addresses the existence of vowel sequences leading to contractions, to rising diphthongs, and even glide insertions.

40 Campbell et al. 1986 invoke this criterion frequently in their assessment of possible features, both phonological and nonphonological, that might identify Meso-America as a linguistic area (sprachbund).

contact is justified is a different issue,[41] but if this principle is valid, the absence of certain sounds also fails as a Balkanism as far as the naturalness criterion is concerned.

5.4.1.4 Relevant Historical Perspectives

Some of the segmental or featural loss must be viewed from an historical perspective and judged accordingly. For instance, the absence of vowel length may well be Balkanologically significant in cases such as Torlak BCMS, inasmuch as it represents a way in which Torlak diverges from other BCMS dialects and it is clear that Torlak is innovative here. In other cases, though, especially for Greek and Romanian, the loss of vowel length happened too early to be Balkanologically relevant, being found already in the Hellenistic period for Greek and in Vulgar Latin for Romanian, thus well before any significant contact among any of the extant Balkan languages.

Similarly, the absence of a rounding opposition in front vowels in Slavic is an inherited characteristic, inasmuch as Common Slavic had no front rounded vowel, but in Greek the situation is different, subject to naturalness concerns (see above, §5.4.1.3), since Ancient Greek had such a vowel (Classical Attic-Ionic Greek < υ > stood for high front round *ü* (IPA [y])).[42]

So also with regard to vowel nasalization, already potentially suspect on the grounds of naturalness (see above, §5.4.1.3, and note also §5.4.1.1, iv), the absence of this feature may not be as significant as some commentators might suggest, once the history of the absence is taken into account. That is, there are some languages that never had distinctive nasalization; Greek is such a language and the same may well hold for Romanian (see Sawicka 2000b). Still, it is interesting that there are Balkan languages that once had vowel nasalization but now lack it: Common Slavic vowel nasalization survived longer in East South Slavic territory (i.e., in Macedonian and Bulgarian) than in the adjacent West South Slavic area, and Common Albanian nasalization is maintained in Geg Albanian, though not in Tosk. But even for those languages there is more to be said. For instance, while denasalization is indeed found throughout Macedonian and Bulgarian, those dialects of Macedonian in the extreme peripheral southwest and southeast of Macedonian-speaking territory, i.e., in Greek Macedonia, retain evidence of nasal vowels (see §5.4.1.1, iv above, but also §5.4.5.1 below; cf. Sawicka 2000b as well), although this evidence may in fact indicate early contact-induced loss of nasal vowels (Friedman 2018a & references therein).

41 Note though that in southern Montenegrin, contact seems to be responsible for the addition of *ü* to the vowel inventory – see §5.4.1.1, regarding (5.9c) – and so too with Ottoman-era Edirne Greek (see §5.4.1.1, iii) .

42 As was the case with *ü* (see §5.4.1.1, iii), so also is it that OCS may have had a front rounded nasal vowel, based on the presence in the Glagolitic alphabet of a special letter so interpreted. If so, it did not last long in South Slavic; see Lunt 2001: 26, 30.

Within Albanian, moreover, while Geg in general still has nasal vowels, there are two interesting exceptions. In the town dialect of Debar in western North Macedonia, in a multilingual (Geg and Macedonian as well as Turkish) urbanized setting, nasalization is lost altogether from the local Geg dialect, as in the local Macedonian,[43] while the Orthodox Geg speakers of rural outlying areas retain nasality on at least one vowel, albeit weakly. Nasalization is also lost completely in the Northeast Geg dialect of the town of Ulqinj (Mtn Ulcinj) and adjacent rural dialects in Montenegro (Gjinari 2007: 76–80; Morozova & Rusakov 2018ab; Morozova 2021), and here, too, a situation similar to that of Debar is a likely explanation. Moreover, in the other direction, since nasal /ã/ (or its denasalized equivalent) tends to be strongly labialized in the relevant northern Albanian dialects (and elsewhere), Stevanović 1935: 20 suggests that the rounding of /a/ in native words in some southern Montenegrin is due to precisely this type of pronunciation in Albanian. Similarly, the Geg of Arbanasi (Itl Borgo Erizzo) in Dalmatia, near Zadar, underwent a loss of nasalization, in the midst of nonnasalizing Slavic (Croatian) speakers.[44]

Thus again one is confronted by the need to recognize contact-based localistic determination in understanding even widespread historical developments.

5.4.1.5 Variable Realization of Features

The realization of some features can be varied even across the different languages that show them. A relevant example is feature (5.9g), vowel reduction (whether phonetic or phonological – see above, §5.4.1.3) in unaccented syllables, with consequent alternations between accented and unaccented forms. Several languages show such effects but in different ways. For instance, in Greek, particularly in northern dialects, the unstressed high vowels of earlier Greek (still preserved as such in southern dialects, including the Standard language) were reduced to zero (i.e., deleted),[45] and subsequently the unstressed mid vowels of earlier Greek (still preserved in southern dialects, including Standard Modern Greek) were raised (i.e., reduced in inherent duration); these developments and subsequent alternations can be illustrated in the following forms (where Ø signals a deletion site) (Table 5.1):

43 In Debar Albanian, Geg *â* yields open [ɔ], the same outcome that Common Slavic **ǫ* has in Debar region Macedonian (namely Reka and Drimkol-Golobrdo).

44 It cannot be argued that Arbanasi Geg never had vowel nasalization, as these speakers relocated there from two villages near Shkodër, a city in a Geg-speaking area with nasal vowels.

45 So also in Judezmo: Afendras 1968: 94 notes that in some Judezmo dialects, e.g., that of Bitola (but not Bucharest – cf. Sala 1971: 100, 108), "unstressed /i, u/ often disappear." However, following Crews 1935: 36, he is reluctant to attribute this development to "borrowing from Greek . . . [as it] reflects changes known in Iberian Spanish dialects as well" (so also in Portuguese) and thus as far as Judezmo is concerned it may be a feature the speakers brought with them to the Balkans.

Table 5.1 *A key Greek vowel isogloss*

Southern (StModGrk)	Northern
περιμένει 'wait.3SG'	πιρØμένØ
περίμενε 'wait.IMPV.SG'	πιρίμινι
άνθρωπος 'man.NOM'	άνθρουπους
ανθρώπου 'man.GEN'	άνθρώπØ

Thus, earlier unstressed *i*/*u* are deleted and earlier unstressed *e*/*o* raise to *i*/*u*, but, given that Greek accent placement can fall on different syllables in different grammatical categories, alternations arise in related forms between zero and stressed *í*/*ú* (representing earlier *i*/*u* and *í*/*ú* respectively), and between *i*/*u* and stressed *é*/*ó* (representing earlier *e*/*o* and *é*/*ó* respectively). Other kinds of plausibly stress-related vowel alternations, e.g., involving a centralized unstressed vowel, do not occur in Greek.[46] In eastern Bulgarian and southeastern Macedonian, reduction occurs but it is most evident with unstressed *a*, somewhat less so with unstressed *o*, and least frequent with unstressed *e*. For the most part, a similar sort of reduction involving *e*/*o* in Albanian is not generally found,[47] though one dialect, that of Kavaja, does attest raising of unstressed *o* (e.g., [gudít] 'hit' for Standard *godit*; Çeliku 1974; Sawicka 1997: 20); far more widespread in Albanian, particularly in the north, is the reduction to zero (deletion) of unstressed *ë* (e.g., [ktu] for Standard *këtu* 'here'; Sawicka 1997: 20). In Romanian, alternations conditioned by stress occur but involve stressed *a* alternating with unstressed *ă*, though the raising of /e/ occurs; as Sawicka 1997: 19 puts it, "in Dako-Rumanian dialects . . . unstressed /e/ is pronounced as very high, or it changes into /i/ nearly on the whole Dako-Rumanian area."

Aromanian shows mid-vowel raising quite regularly, affecting both /e/ and /o/, e.g., in Kruševo (Gołąb 1984a) one finds *véd*^{u} 'I see' but *vidém*^{u} 'we see,' *pót*^{u} 'I can' but *putém*^{u} 'we can,' etc. The Frasheriote Aromanian of Ohrid-Struga has taken this process even farther, reducing, i.e., by raising, final unstressed /e/ in general, e.g.,

46 Interestingly, the Greek of the transhumant Karakačan shepherds in Bulgaria has unstressed /a/ > *ə* (a change otherwise unknown in Greek), but this could be the result of contact with dialectal Balkan Slavic (thus another very localized phonological Balkanism). This occurs only in Bulgaria; note too that unstressed /a/ sometimes becomes *e* (Høeg 1925: 145–146) and even *i* before or after a liquid: *valaniðiá* > *vilaniðiá* 'oaktree' (p. 154), and also stressed /e/ > *æ* when there is a back vowel in the following syllable (158–159) – Skok 1927 compares this to Romanian.

47 Thomason 2001: 108 refers to vowel raising in the Balkans as "the raising of *o* to *u* in unstressed syllables in Bulgarian, Rumanian, and Albanian," echoing the language distribution given for this feature in Miklosich 1862: 8 and so with no mention of the change of *e* to *i*. The basis for including Albanian at all, however, is unclear to us; Miklosich cites Hahn's1854 grammar but Hahn (p. 13) lists only four lexical items, of which three are not labeled as to dialect (glosses ours): *koronë* = *kuronë* 'crown' (this form is Geg; Tosk would be *kurorë*), *molitë* = *mulitse* 'moth,' and *rrotulloj* – *rrutulloj* 'I encircle' (one would expect *rrotullonj* for most of Tosk), and one is labeled Tosk *orji'* = Geg *urji'* 'anger' (*sic*; StAlb *uri* means 'hunger' and there is a colloquial word *urejtje* 'hatred').Thus there is no real indication of anything systematic about the Albanian facts regarding apparent vowel raising. It is surprising, moreover, given that this is a prominent feature of Northern Greek, that no mention is made of Greek with regard to this feature (though other corrections and additions are needed, i.e., to include Macedonian and Aromanian in the discussion).

Kruševo *múnte/múntile* Frasheriote *múnti/múntili* 'mountain.INDEF/DEF, Kruševo *neádze* Frasheriote *nedzi* 'go.3sg.PRS (see Markovikj 2007: 27ff. on this and other reductions). Moreover, final unstressed high vowels are reduced to nonsyllabics in VC_# or V_#, and in South Aromanian final unstressed /e/ raises to [i] when the syllable is open (Gołąb 1984a: 33–34).

Finally, some dialects of Judezmo, e.g., that of Bitola, show raising of /e, o/ to [i, u] (Afendras 1968: 94, drawing on Crews 1935), but only in word-final position, unlike the other Balkan languages showing raising; moreover /a/ is also subject to raising, to /e/, again unlike any other Balkan version of vowel raising. However, vowel raising in the Judezmo spoken in Bulgaria today does show influence from Bulgarian vowel raising (Grünke et al. 2023). Putting all this together,[48] it is clear that while a common core for some aspects of vowel reduction, in particular just the vowel raising, can be identified in a geographically specific range, trying to subsume all vowel reduction phenomena in the Balkans under a single rubric cannot be accomplished. It may well be that some subsets of these developments could be regionally defined traits, i.e., local Balkanisms, and as a result this feature is revisited below in §5.4.3.9.

5.4.1.6 Possible Contact Bases

Most significantly, if, as argued in §3.5, our definition of "Balkanism" rests on contact-induced features as opposed to merely accidentally convergent ones, then one must ask whether the abovementioned phonological features could legitimately be considered as resulting from language contact, especially in light of the mechanisms for phonological contact effects identified in §5.2.

As indicated in the previous section, so too in the case of the loss of various features in the vowel system, such as nasalization or rounding or length, the changes could in principle be contact-induced if it is assumed that these are extra sorts of features that, while potentially distinctive, nonetheless go beyond basic qualitative differentiation in a set of vowels. As such, especially if any of these features are absent in the L2 learner's native language – note, for instance, that length distinctions had disappeared in Greek and Latin at a relatively early period (see §5.4.1.4) – they might be particularly susceptible to reduction and leveling out in a second language acquisition scenario, thus setting the stage for mutual simplification, the lowest common denominator effect referred to above (in §5.2).

With other vowel changes, however, it is hard to make a case that they could be contact-induced developments since native developments can be involved. For instance, in the case of the schwa feature, there are loanwords that show schwa in what were surely originally stressed positions, e.g., Albanian *këndoj* 'sing,' with stress on *-o-* but from Latin *cantō*, with stress on *-a-*, and loanwords that show schwa and

48 The situation is more complex, and a single treatment more difficult to attain, than this brief account might suggest, as it leaves out a large number of reductive types of effects in unstressed syllables in which different vowels are affected or different outcomes are found (e.g., for widely attested *çorap* 'stocking' in Albanian, the dialect of Hot has *çarap*); see Sawicka 1997: 18–23 for a succinct but detailed presentation of the relevant facts.

maintain a stress, e.g., Geg *kângë* from Latin *canticum*, for which Tosk has stressed schwa *këngë.* However, there are also native developments in which regular sound changes yield a schwa, e.g., in Tosk Albanian, from earlier nasal vowels, e.g., *është* 'is' = Geg *âsht*, or from PIE *-om as in Geg *kê* = Tosk *kë* 'this.ACC.M' ultimately from *$k^{w}om$. Thus, the stressed schwa in Tosk is a specifically Tosk development connected to the Geg–Tosk split, not directly contact-induced, with words of Latin origin being treated just like native words. Moreover, in some instances, there are several different sources for the schwa in each of the relevant languages.[49] Thus, one would have to assume that words with schwa were borrowed from one language (that had schwa) into another that did not have schwa and that speakers of the recipient language, for whatever socially determined reason, chose to adopt such forms without altering the schwa to some sound that was, for them, native. Such developments do happen, as the examples involving Aromanian in §5.3 indicate, but are perhaps rarer than alteration of loans. Moreover, one would then have to assume as well that the schwa made its way into native vocabulary in the borrowing language. Again, that is possible and in fact such an example involving Aromanian is given above in §5.3, but it can generally be considered a less usual outcome.[50]

What becomes harder to see as contact-induced, however, is the fact that the schwa in native vocabulary results from several earlier sounds in some of the languages; for instance, the Romanian mid central vowel *ă* derives when stressed from Latin *o* in some contexts, as in *contra* 'against' > *către* or *foras* 'outside' > *fără*,[51] and the high central vowel *î*, when stressed, derives from (stressed) *a* before a nasal, as in *câmp* 'field' from Latin *campus*, and from *i* after *r*, as in *râpă* 'cliff' from Latin *ripa*. Moreover, in each of the languages with stressed schwa, the developments that created it have the appearance of regular sound changes, in that the changes affect a number of lexical items without exceptions. Thus under a contact-induced account of the emergence of stressed schwa in the Balkan languages, one would have to assume that the mere occurrence of a few stressed schwas in loanwords was sufficient to lead different sounds in each of the recipient languages in the direction of schwa, in such a way as to appear to be a regular, exceptionless sound change.[52] Moreover, if one were to posit that interference or

49 Thus, for example, the northern dialects of Macedonian (e.g., Tetovo, Kumanovo, Kriva Palanka) have schwa from Common Slavic jers, eastern central dialects (e.g., Tikveš-Mariovo) have schwa from vocalic *l,* many peripheral dialects (e.g., Ohrid-Prespa) have schwa from the Common Slavic back nasalized vowel, and so on. Friedman 1993d: 301 has a convenient summary of the facts. It is worth noting that Turkisms in *ıl, ır* give the same reflexes as inherited vocalic /r/ and /l/, e.g., Trk *kaldırma* > western central and StMac *kaldrma* 'cobblestone,' *asıl* > *asolen* 'genuine.' See §5.4.1.1 on sources of stressed schwa in some Geg dialects.

50 A similar example is found in some western central Macedonian dialects, where the Turkish high back unrounded vowel is sometimes borrowed as such or as schwa, as in *kăsmet* 'fate' (from Turkish; although expected *kasmet* in western central Macedonian also occurs), but the sound also occurs in *găz* 'backside,' which is native but comes from a different dialect since the schwa here (*ă*) is a reflex of a back nasal.

51 Moreover, adding unstressed contexts into the mix only further complicates the picture, as both *a* and *i* can, when unstressed, yield *ă*.

52 Although this makes it hard to accept stressed schwa as a Balkanism, we do note that Blevins 2017: 98, in propounding her Areal Sound Pattern Hypothesis, does say the evolution of areal sound patterns "may mimic that of internally phonetically based sound change."

substratum effects were the cause and not diffusion from lexical borrowings, one would be faced with explaining how speakers of a language with a schwa carried over their schwa-producing habits into their pronunciation of the words of another language, including in stressed positions where vowel reduction in the direction of a schwa is less expected. And even if this were plausible, it is hard to link the developments in one language, say Albanian, with those in another, say Bulgarian, where different sets of L2 speakers would be involved, and where, moreover, we lack adequate evidence concerning substratum languages.

Thus despite the insistence over the years that the overall schwa developments could well be a significant shared feature in the Balkans, it is not at all clear that they have anything to do with language contact, other than occurring in some loanwords (so Petrucci 1999: 62–69 regarding Romanian, explicitly countering claims of Slavic structural influence by, e.g., Rosetti 1958, 1964a; Petrovici 1958; Shevelov 1965, and others following them). Cf. Hamp 1977a: 281–282, who explicitly rejects the characterization of schwa as a Balkanism. Sawicka 2000b, however (see also Sawicka 2000a), considers accented schwa to be at least a Medieval Balkanism. The key question in trying to treat schwa-emergence as a once-active process by which schwas were innovatively created at a time when schwa was not a part of the inventory of these languages is this: what would the relevant creative mechanism be in this case? One has to wonder as well whether sounds by themselves can be borrowed, and whether the habits of pronunciation in a target language can be replicated by speakers of another language and adopted into their native language. Ramos-Pellicia 2004 reports on some cases of speakers of Puerto Rican Spanish in South Lorain (Ohio) pronouncing Spanish words with American English bunched [r] rather than the expected vibrant trill more typical of Spanish; and, interestingly, this even occurred in speakers without much exposure to or knowledge of English. Still, even in that case, she is more inclined ultimately to attribute it to the presence of Chicano Spanish speakers in Lorain who seem to have acquired an English-type [r] prior to arriving in Lorain as the result of contact with English (e.g., in the southwest of the United States) and thus it comes only indirectly from English into Lorain Puerto Rican Spanish; moreover, the production of American [r] by these Puerto Rican Spanish speakers was quite sporadic (and restricted to word-list production, not spontaneous speech) and may even have been used by some speakers as a device for avoiding the stigmatized (for Puerto Rico) [l] for /r/ in final position. Thus even if borrowing of sounds per se is possible, it is not obvious how that alone could lead to the widespread, regular, and systematic occurrence of a particular sound in native vocabulary in the various languages in question.[53]

53 There are cases, such as retroflexion in Indic languages or clicks in Southern Bantu languages, where borrowing is fairly clear as a, if not the, significant factor in the occurrence of these sounds. In the case of Indic retroflexion, at least, as argued by Hock 1975, there were language-internal sources of retroflexion that may have played a role, so that outside influence need not be the sole cause. As for the Bantu situation, it seems fairly clear that the institution of *hlonipha* (see Gal & Irvine 1995; Irvine & Gal 2000) in Nguni languages, a form of avoidance speech, provided a motivation for the

What we would say is possible, however, is that on a local level, in terms of more narrowly defined clusters of languages in direct contact in particular regions, schwa could be a Balkanism. This is taken up briefly in §5.4.5.1 below.

5.4.1.7 Summation

In general, looking to vowel correspondences for something Balkanologically interesting other than at the local level is difficult (and, we would say, misguided) because even when there is a common element, it can have multiple sources in each language, and often is simply the result of regular sound change. However, we would argue further that even if one insisted that there is something superficially significant from a Balkanological perspective regarding vowel systems, it is important to note that such a distribution can be achieved through a localistic approach anyway with clusters of regionally defined features adding up to a broader overall distribution. This idea is revisited below in §5.4.3.

5.4.2 (Nearly) Pan-Balkan Features: Consonants

With regard to consonants, the following have been proposed (5.10), though it is interesting to note that there is less widespread agreement on consonantal "Balkanisms" than there has been on the vocalic features discussed in §5.4.1:

(5.10) Convergences involving consonants
- a. Weakening/lenition (especially of voiced stops) [HS, EB, IS]
- b. Palatalization of velars (and sometimes dentals), especially before front vowels, thus giving hard and soft velars (or dorso-palatals or palatals) [EA, EB; JF, IS]
- c. Many common consonant phonemes in all languages: p/b, t/d, k/g, f/v, s/z, m, n, l, r, j [EA, JF]
- d. "Hissing" / "hushing" opposition (roughly: apico-dental / alveo-palatal) in fricatives and affricates, i.e.: s/š, ts/č, and so also for voiced [JF]
- e. Presence of at least two members of the set *c č ć* (using a Slavistically based notation) [JF (citing Ivić 1968), discussion in IS]
- f. /n/ – > [ŋ] /__k/g (allophonically) [JF]
- g. Presence of *h* (*x* in some languages), with fronted allophones before front vowels [HS 131, JF, PA]
- h. Rarity of geminate consonants (except across morpheme boundaries) [JF, IS]

Some of the same sorts of issues that were raised with regard to the putative pan-Balkan vowel convergences can be raised with regard to the consonant features, pertaining to distribution, naturalness, chronology, and the like. These are taken up in turn in the subsections that follow.

entry of the ingressive velaric consonants that are more common and complex in the neighboring Khoisan languages. In the Balkan case, we see this latter type scenario as unlikely.

5.4.2.1 Distribution of Balkan Consonant Features

As with the putative pan-Balkan vowel features, the consonant features that have figured in various discussions are not found in all the relevant languages and thus show a scattered distribution. Still, some of features do reveal what might be Balkanologically interesting geographic distribution. For instance, among Greek dialects, it is the southeastern Aegean dialects, including Rhodian but especially Cypriot, i.e., those most removed geographically from fully Balkan varieties of Greek, in which geminates are a prominent part of the phonology, so that the absence of geminates from the rest of Greek is in keeping with (5.10h). The other languages also lack geminates, except for Bulgarian, though they occur at morpheme boundaries and have arisen secondarily here and there, including in some mainland dialects of Greek (so Newton 1972; Sawicka 1997: 36).

Similarly, the more central Balkan northern dialects of Greek have the widely distributed (i.e., found in Albanian, Balkan Slavic, Balkan Romance, Romani, Judezmo (in Bucharest, Sala 1971)) hissing/hushing opposition of (5.10d) and the multiple affricates (specifically *c č*) of (5.10e).[54] Thus these all have a more Balkan-like phonology in these regards than what is found in Standard Modern Greek (based as it is on a southern dialect) with only dental τσ ([t^s]) and τζ ([d^z]) affricates and no others,[55] or the geographically peripheral standard Turkish, which also distinguishes only one point of articulation for affricates, having just palatals and nothing else.[56] This mismatch between Greek and Turkish has meant that each language substitutes its class of sounds for the other's in adapting loanwords from the other language. Importantly, though, West Rumelian Turkish tends to have both palatals and dentals, thus diverging from standard Turkish in the direction of co-territorial Balkan phonological systems. Judezmo likewise diverges from other Spanish dialects with its *c*, *dz*, and *dž* (*č* occurs in many Spanish dialects besides Judezmo; see Quintana Rodríguez 2006: 82–84 on distributions in Judezmo).

54 Cf. Papadamou & Papanastassiou 2013: 393. Note also that in some of the Greek dialects of Aegean (Greek) Macedonia *sk* > *šč* before front vowels, e.g., *skivo* > *ščivu* (Papadamou & Papanastassiou 2013: 395), which, despite its current spotty distribution in northern Greek, looks very much like a Slavic substratum feature; see also §5.4.4.5.

55 We label Greek τσ as an affricate here, recognizing the controversy in Greek – one of the "dreams of modern Greek phonology" discussed in Householder 1964 – as to whether it and its voiced counterpart τζ are true affricates or instead consonant clusters. Joseph & Philippaki-Warburton 1987: 231–232, 238 offers some discussion of the analytical ambiguities and Joseph & Tserdanelis 2006 (see §5.4.6 below) discusses the relevant phonetic properties of the Greek voiceless sound; see footnote 124 regarding Greek dialects that contrast cluster [ts] with affricate [t^s]. The literature on comparable Macedonian sounds is similarly controversial: Hendriks 1976: 59–60 talks about *c* and *dz* as "phonetic … close-knit affricates" but argues on the grounds of "distinctiveness" for a "biphonematic [i.e., cluster] interpretation"; Lunt 1978: 138 criticizes the biphonematic analysis, as it forces "the consonants … into the Dutch pattern … because of a 'distinctiveness principle' that, in the 1970s, seems … merely bizarre."

56 We should also note that in the case of Romani, the opposition *c*/*č* is limited mostly to loanwords, and it is lacking in the Balkan II (Burgudži) dialects, where alveo-palatal and dorso-palatal affricates become dentals. Among Romani speakers, this feature is commonly believed to have been acquired in Greece, after which the speakers migrated northward (VAF field notes), although in fact the dialects appear to have originated in northeastern Bulgaria and then spread westward as far as Kosovo.

Finally, in the socially peripheral Romani, affricates other than the palatals are not native to the common (pan-Romani) system, but are found just in loanwords and in various dialects under certain specifiable conditions (including as a regular sound change). In fact, Romani dialects that have been involved in intensive contact with either Greek or Turkish generally show the same types of affricate adaptation in loanwords (and even native words, cf. Igla 1996: 9; Cech & Heinschink 1999: 10; Friedman 1988a) as occur in the contact language.

Still, the opposition noted in (5.10d) between, on the one hand, mellow alveo-palatal or palatovelar stops or affricates or palatalized velars, and, on the other, strident palatal affricates is found across much of the Balkans, though neutralization of this opposition occurs in the local varieties of Albanian, Balkan Slavic (parts of Serbia, Kosovo, and North Macedonia), WRT, and Romani in Kosovo and Macedonia (Ajeti 1998: 101–103; Hafiz 1979; Friedman 2017a).[57] On the other hand, the palatalization or palatal affrication of velars in Judezmo occurs precisely in contact with Slavic and Romanian (Quintana Rodríguez 2006: 92–93, 380).

The matter of affricates in the central Balkans is taken up again, in a different context, below in §5.4.4.2, where some additional facts about their distribution are presented.

Moreover, others of these features are rather limited in their extent. For example, the lenition of voiced stops to fricatives essentially holds just for Greek, a development that began early in the Koine period, to judge from occasional misspellings in inscriptions and papyri. However, Sawicka 1997: 47 quite reasonably suggests that the loss of intervocalic (more properly medial, i.e., word-internal) voiced stops in Albanian, a change which affects Latin loanwords (e.g., *mjek* 'doctor' from *medicus*) as well as inherited words from Proto-Indo-European (e.g., *det* 'sea' from **dubeto-*, lit., 'the deep (thing)'), can be viewed as a weakening of these stops for this language too.[58] She further doubts (p. 48) that lenition is even "characteristic for the Balkans," based in large part on this limited distribution. There is as well the loss of intervocalic voiced consonants in Macedonian dialects (cf. Koneski & Vidoevski 1983: 128–131), but this seems to be relatively recent and thus independent of other Balkan lenitions.

5.4.2.2 Definitional Issues

The consonant features in (5.10) do not present any particular issues with regard to definition, in the way that diphthongs did for the vowel features. It is interesting, however, that while several scholars (JF, HS, PA) point to the presence of *h* or *x* as a widespread Balkan feature found in most of the languages (though its occurrence in Romanian is often attributed to Slavic influence, e.g., through loans, so Petrovici

57 The distribution of mellow/strident neutralization for Albanian is complex and includes some parts of western North Macedonia, Kosovo, and northern Albanian (Gjinari 2007: 94–95). The place of articulation of mellow palatals in Albanian and Balkan Slavic tends to be the same in regions where the two are co-territorial: more forward to the north and further back to the south.

58 See also §5.4.4.10.2 for more discussion of the loss of intervocalic voiced stops.

1957: 41; Schaller 1975: 131; and see also Petrucci 1999: 83–84), the opposite is more Balkanologically relevant. Sawicka 1997: 34–35 notes the loss of [x] as an important feature (see below §5.4.4.6). We add to this the fact that in western Macedonian [x] is completely eliminated or changes to [v] or [f] (depending on position and additional factors), and the change to [f] (and sometimes elimination of /x/) also occurs in adjacent or co-territorial Albanian dialects. Such eliminations also occur in Romani and WRT. Clearly different aspects of the phonology are being focused on in order for both presence and absence to be held to be significant, but this inconsistency demonstrates the importance of having clear criteria for judging particular effects. In this regard, it is worth noting that some dialects of Romani oppose [h]/[x], contrasting, e.g., *hasav* 'I laugh' with *xasav* 'I cough,' so that the terms of the feature would have to be judged differently for those speakers (see Boretzky & Igla 2004: Map 23 for details).

5.4.2.3 Balkan Consonant Features and Naturalness

Particularly vexing, though, for several of the consonant features in (5.10) is the matter of naturalness: to the extent that a given feature is cross-linguistically quite common or phonetically rather trivial, its presence in any language – in the Balkans or elsewhere – is, as pointed out in §5.4.1.3, not particularly noteworthy. Phonetic naturalness might offer a rationale for at least one feature, namely (5.10h), referring to the rarity of geminates, in that the general absence of geminates in the Balkans, for at least some of the attested languages, is the result of a loss. Such an elimination or absence would accord with the sensibility that geminates are more "marked" cross-linguistically and require greater articulatory effort, and are therefore more likely to be eliminated in a contact situation. However, even if it provides a motivation for the feature, naturalness also undermines the potential significance of this feature as a Balkanism, since the absence of geminates from the Balkans is perhaps no more striking than their absence from any non-Balkan language.

Several other of the consonantal features cited fall short of being compelling as potential Balkanisms when measured against this naturalness criterion. Clearly, a characteristic such as (5.10c), referring to common consonant phonemes in all languages (*p*/*b*, *t*/*d*, *k*/*g*, *f*/*v*, *s*/*z*, *m*, *n*, *l*, *r*, *j*) is hardly meaningful as a likely contact-induced trait since these sounds are among the most common cross-linguistically. This criticism was already made by Sawicka 1997: 31: "It is not difficult to see that this part of [*sic*] consonant system the Balkanic languages have in common is trivial. It is the same or nearly the same in a number of other European languages." Indeed, languages outside of Europe also have many of these sounds, e.g., Arabic, Cree, Hebrew, and Javanese, to name just a few, and Maddieson 1984: 32 notes that over 99 percent of the 317 languages surveyed have, e.g., labial, coronal, or velar stops. Similarly, the palatalization of velars before front vowels (feature (5.10b)), even if it results in a systematic set of soft consonants alongside hard consonants, is an extremely common phonetic process, found at least allophonically in numerous languages all over the world.

One aspect of palatalization, however, may be worthy of mention and is instructive as it shows what sort of situation could be Balkanologically significant even in the face of naturalness problems. In particular, Sasse 1991: 61 notes that Arvanitika shows the Greek type of allophony with /x/, in that it is realized as the palatal fricative [ç] before front vowels. This is a phonetically natural development to be sure, but if, as it seems, it is restricted just to Arvanitika within Albanian, then, given that most speakers of Arvanitika are bilingual in Greek, the fact that it mirrors a Greek distribution makes it look like a reasonable candidate for a contact-induced feature. Moreover, it would in this way be in keeping with the localistic thrust in the approach to Balkan phonology advocated here.

Finally, the assimilation that Feuillet points to whereby /n/ is realized as velar [ŋ] before velars reflects a rather trivial conditioned articulatory adjustment. However, it is true that within Slavic at least, there are languages in which an *-nk-* sequence does not show assimilation in at least careful speech, though assimilated pronunciations do occur; Ukrainian family names in *-enko*, for instance, are typically pronounced [. . . enko], though the assimilated form [. . . eŋko] can be heard in casual or fast speech, and Schenker 1980: 209 reports that where Warsaw Polish has only [nk], pronunciations with [ŋk] occur in the Polish of Cracow and Poznań. Thus the more frequent and robust assimilation of /n/ to [ŋ] in Balkan Slavic may be a significant enough divergence from the pan-Slavic norm to be of potential interest from a language-contact perspective. Still, the naturalness of this particular development means that it could easily arise in any language without the impetus of language contact (as the fast speech pronunciations in Ukrainian might suggest).

Thus, overall, there is good reason, on the grounds of naturalness, for doubting the probative value of these features as defining characteristics for the sprachbund.

5.4.2.4 Relevant Historical Perspectives

These last two examples in §5.4.2.3, however, show that for the consonants, too, part of what is potentially significant is not just possible convergences among neighboring languages but also divergence away from earlier states. Often, as in the above two cases, the divergence needs to be measured by comparison with related languages not in the Balkans, i.e., by judicious application of the traditional (and effective) comparative methodology of historical linguistics. Nonetheless, many of the claims embodied in (5.10) can be evaluated as problematic (or not) even without much comparative evidence needing to be brought into play.

5.4.2.5 Variable Realization of Features

The consonants that figure in the list in (5.10) are not particularly subject to variation in their realization, unlike some of the vowel features. Still, it can be noted that /h/ and /x/, treated as if they were mere variants of one another in (5.10g), do represent distinct sounds and in fact, as noted in §5.4.2.2 above, some Romani dialects have both sounds in contrastive environments.

5.4.2.6 Possible Contact Bases

For the most part, the features listed in (5.10) are ones that are conceivably brought about by some type of language contact, the exception being (5.10c). Individual sounds may enter a language via loanwords, and indeed the likely source of (most instances of) Romanian /h/ was via unadapted loanwords from Slavic[59] at least in the earliest period.[60] Native sounds may undergo adjustment in the light of entering foreign sounds, e.g., Hamp 1989b: 200 argues that inherited *h* in Arvanitika, once *γ* entered the language through Greek loanwords, "moves opposite /γ/, taking on a velar spirant articulation." Native language pronunciation habits often carry over via transfer into the phonology of a secondarily acquired language, as noted earlier, so that, for example, assimilating a nasal next to a velar becomes plausible if the speaker has been doing it automatically in his/her native language. The same can be said of palatalizing consonants before front vowels. However, it is hard to see how having, or coming to have, a particular set of fifteen phonemes, as in (5.10c), possibly augmented with *h*/*x* from (5.10g), could spread from one language to another, except possibly through numerous mergers and repeated losses of distinctions induced by accommodation in face-to-face communication, the lowest common denominator effect referred to earlier; but even then one has to wonder why the same mergers and losses would have occurred in each set of languages in contact.[61] Similarly, it may not be reasonable to suppose speakers could somehow choose from among a set of sounds, as a perhaps uncharitable interpretation of (5.10e) would suggest.

5.4.2.7 Summation

Overall, even though individual items among the consonant features may have some plausibility along one or more dimensions, there are enough problems associated with each of the items in (5.10) to raise cautionary flags about adopting them uncritically as significant Balkan features. Moreover, there is a key issue, as has already been suggested, that holds for all putative Balkanisms, vowels and consonants alike, namely how to account for and achieve spread over a large area? Such problems lead to the need to look at phonological features on a localized

59 For instance, Late Common Slavic **duxŭ* > Rmn *duh* 'spirit,' **xrana* > Rmn *hrană* 'food,' and **xuliti* > Rmn *huli* 'to slander.' There are also words in which Slavic **x* was replaced by [f] in Romanian, as in **ruxo* > *rufă* 'linen,' likely reflecting an earlier layer of borrowings, perhaps before Slavic was more familiar to early Balkan Romance speakers. See Petrucci 1999: 82–86 for some discussion.

60 The Slavic source of Romanian *h* is disputed by Brîncus 1961, though Petrucci 1999: 83–84 convincingly counters his arguments, which never seem to have won much support.

61 One answer might be markedness (cf. Trudgill 2004), with the unmarked member of a set being selected repeatedly and independently because it is unmarked. We are skeptical here, in the absence of clear criteria allowing an individual speaker, faced with just ambient data from one or even a few languages, to know what is marked and what unmarked. The only overt basis a speaker could have is the frequency of particular sounds, but this should vary from language to language and thus offer no cross-linguistic basis for deciding markedness.

basis; several such localistically determined features have already been proposed but more about each one, and possibly others, can be said.

5.4.3 Regional Features: Vowels[62]

Accordingly, we turn now to local Balkanisms pertaining to vowels that have been proposed in the literature. So as not to belabor points made earlier about other features even if relevant here, we list each one in its own subsection with some pertinent discussion. While it should be clear that we are generally sympathetic to the idea of looking to regional features, not all that have been proposed are equally compelling or equally well considered in the literature; thus various features cannot simply be accepted outright without an appropriate degree of scrutiny. In the section heading for each feature we discuss, we list the languages that possibly show parallelism. The matches across languages that have been discussed most in the literature or are, in our estimation, more compelling are separated by commas, while the less compelling matches are separated off by a semi-colon (thus, e.g., in §5.4.3.1, the listing "Grk, Aro; Jud" means that the parallelism between Greek and Aromanian deserves attention or has gotten particular consideration by scholars as a possible local Balkanism, but adding Judezmo for this feature is more speculative).

5.4.3.1 *au* > *av/af* (Grk, Aro; Jud) [KS102; EA95-6]

This feature, first discussed by Sandfeld, involving the increased consonantalization of the offglide of the *au* diphthong,[63] with the *v/f* difference being governed by the voicing of the following segment, is indeed found within the Balkans in Greek and Aromanian, as in AGrk παύω 'I stop' / παῦσις 'pause' (both with [au̯]) > ModGrk παύω ([pávo]) / παύση ([páfsi]), and Latin *laudō* 'I praise' (also with [au̯]) > Aro *alavdu*. This change in itself is not a particularly unusual one, being found, as Sandfeld notes (p. 102 n. 1) in Russian (e.g., [*zaftra*] 'tomorrow,' orthographic *zavtra*, from **zawtra* (from **za utra*)) but also in Ukrainian, among other languages, and it was a relatively early development in Greek, with traces evident in the Koine period. Also, in Greek it has a wider scope, affecting not only Ancient Greek αυ but also ευ.[64] Nonetheless it could well be a feature that Aromanian

62 We are thus omitting here, relegating it to footnote status only, the mention by Sandfeld 1930: 146 of the development of initial syllabic /r/ and vowel epenthesis in final clusters in both (Serbo-)Croatian and Istro-Romanian, since these languages lie at the outer edges of our geographic focus. We suspect language contact is involved in this development in Istro-Romanian, removed though it is from the territory of Balkan languages proper.

63 It thus also invites inclusion with the consonantal regional features.

64 This change also affects the outcome of the Ancient Greek long diphthong ηυ, giving Modern Greek ηυ ([iv/if]); this sequence is of very limited occurrence, being restricted in the standard language to a few high-style (Katharevousa) past tenses (e.g., διηύθυνα ([diífθina]) 'I directed' from earlier διηυθυν-), though it is of wider occurrence in some dialects (e.g., ηύρανε ([ívrane]) 'they found' in the Greek of southern Albania, from earlier ηὗρον, and even among some speakers of the standard language).

speakers have gotten from Greek due to bilingualism in Greek, especially since it represents a divergence from Romanian (cf. Rmn *caut* 'I seek' vs. Aro *caftu*). Here Meglenoromanian *caft* (Capidan 1935: s.v.) shares the change with Aromanian, and given that in later centuries Macedonian was the main contact language, this looks like an old South Danubian Balkan Romance (SDBR) development, perhaps from before the arrival of the Slavs.

Afendras 1968 (apparently drawing on Crews 1935) adds the Judezmo of Thessaloniki to the languages showing this development, saying (p. 96) that "the second element of a /V u/ diphthong [has become] a /v/"; while he states this development in very general terms, his only example is *sivdað* 'city.' However, there is good reason to doubt the validity of this addition of Judezmo. While Castilian Spanish has *ciudad* for 'city,' taking the Latin source, *CIVĪTATE(M)*, into account indicates that the *-u-* of Castilian is an innovation and that the *-v-* of Judezmo reflects the Medieval Spanish pronunciation, and the same could be said about other examples one might try to adduce here, e.g., Lat *DEBĬTA(M)* > Jud *devda* (vs. Castilian *deuda*). Thus Judezmo is to be discounted here, but nonetheless, the Greek and Aromanian convergence marks this as a very plausible local Balkanism.

5.4.3.2 Ø > *a* /#__C (Grk, Aro; Jud; Rmi) [KS103, PA, IS; EA95-6]

This development involves the addition of a (prothetic) vowel in initial position before a consonant. Formulated as it is here, there seems to be an interesting parallel between the two languages, Greek and Aromanian, said to show this effect. In addition, examples with a prothetic vowel can be found in Romani, and, according to EA (citing Crews 1935: 38) in some dialects of Judezmo, specifically Bitola (Manastir) and Thessaloniki, where he says a "tendency to develop word initial *a* is also operative." In fact, though, the consonants involved and the extent of the prothesis are different for each language. For the most part, the occurrence of the prothetic vowel is rather sporadic in Greek; αβδέλλα 'leech' from AGrk βδέλλα is a widely cited example, though there are others, such as αλησμονώ 'forget' from earlier λησμονώ, ακρυφά 'secretly' from earlier κρυφά (see Thumb 1912: 14 for more), and in Romani, for which examples such as *nav* ~ *anav* 'name,' *bijav* ~ *abijav* 'wedding,' *šunel* ~ *ašunel* 'hear' can be cited, where the initial *#a-* is innovative. The distribution of this feature in Romani appears to have the Vlax dialects (i.e., the territory of Wallachia and Moldavia) as its center, having moved from there only into some eastern Balkan dialects and southern central dialects (cf. Matras 2002: 227–228). The occurrence in Judezmo is also somewhat limited: Crews cites as a form with "*a* prothétique" ('prothetic *a*') *amañana* 'tomorrow' (cf. Castilian Spanish *mañana*), but there are many nasal- and sonorant-initial words in Crews's vocabulary. According to EA, "Crews mentions common *a-* prefixes for verbs" but it is not clear from Crews whether this is phonological prothesis or a true prefix and in any case, the two verbs cited from Thessaloniki have rather different consonantal onsets (*asumbir* 'go up' and *awgumitar* 'vomit'). Moreover, similar

forms occur elsewhere in Ibero-Romance, as both Crews and EA note: Portuguese has *amanha* for 'tomorrow,' for instance. To this we can add that Quintana Rodríguez 2006 does not adduce this feature at all. Thus in Judezmo, the prothetic *a-* may be a Spanish or Portuguese dialectism brought to the Balkans. By contrast to these languages, in Aromanian the consonant involved is generally a liquid and the result is quite regular, to the extent that there are virtually no words with an initial liquid. It must be added that consonantal liquids, especially /r/, are known to cause some articulatory problems in initial position for which vocalic onsets might be a solution.[65] These various considerations, together with the mismatches between the languages in the consonants, suggest that this apparent parallelism, despite the appeal of a Greek-Aromanian (-Judezmo) convergence, is nothing more than apparent.[66]

5.4.3.3 Unstressed Initial Vowel Loss (Slv, Grk) [Seliščev 1925: 49; IS]

An interesting counterpoint to §5.4.3.2 is the suggestion by Seliščev, citing Conev,[67] that the loss of unstressed initial vowels in the Slavic spoken in areas to the east of Thessaloniki in the nineteenth century had its basis in a parallel development in Greek (cf. also Sawicka & Cychnerska 2018: 93–102). The Greek phenomenon is a well-documented and relatively early sound change that occurred on the way to Medieval and Modern Greek and is found all over the Greek-speaking world. It regularly affected all vowels other than #α- and more sporadically affected even α-, as in μέρα 'day' from earlier ἡμέρα, λίγο- 'little' from ὀλίγο-, δεν 'not' from οὐδέν, γελάδα 'cow' from earlier αγελάδα (ultimately from AGrk ἀγελαῖος 'in herds, gregarious'), etc. As for Slavic, although Seliščev gives no details, Conev 1934: 428–429 cites some Balkan Slavic examples of sporadic loss of initial vowels, specifically *a*, *e*, *i*, in words of foreign (Greek) origin, e.g., *[a]natema*, *[e]vangelie*, *[i]kuna* (from *ikona*); while these words could reflect the loss of the vowel in the Greek source, he also cites the sporadic loss of initial *o* in function words, e.g., *[o]vaka* 'in this manner'; *vaka* is standard in Macedonian and considered a characteristic expression. On the other hand Mac *olku* 'this much' < **ovolku* shows the loss of the entire first syllable.

In the absence of details on Slavic, the purported parallelism cannot be verified but what little there is not convincing. Still, the Greek sound change would have meant that in vernacular Greek – abstracting away from the occurrence of any learnèd borrowings

65 Once one recognizes laryngeals as the source for some Ancient Greek so-called "prothetic vowels" (see Beekes 1969 for discussion), Proto-Indo-European becomes a language with few, if any, instances of words with initial *r-*, perhaps reflecting some articulatory "discomfort."

66 Some of the Greek examples cited by Sawicka 1997: 26 as showing prothesis instead show loss of an original unstressed vowel; thus in νωρίς ~ ενωρίς 'early,' the vowel-initial form reflects the earlier state and the vowelless form is the innovation, and so too, most likely, in ρωτώ ~ αρωτώ 'to ask' (if the *a*-vocalism reflects sporadic coloring (especially around a liquid) of an original ε- from AGrk ἐρωτῶ rather than prothesis onto innovative vowelless ρωτώ).

67 Seliščev does not give a specific work of Conev's as his source; since Conev 1934, 1937, 1940, which we had access to, reprints his earlier work, we are confident that the 1934 material reflects what Seliščev was referring to.

with initial unstressed vowels that reflect Ancient Greek forms more directly – it would plausibly have been a phonotactic (or even morpheme structure) condition banning unstressed vowels in word-initial position. Then the only vowels that could occur word-initially would have been stressed vowels, i.e., words must begin with a consonant or a stressed vowel. Such a constraint in one language could well have made its way into and thus affected another language through bilinguals, especially Greek-dominant speakers of Slavic. Even so, it must be admitted that the loss of unstressed vowels in any position is not a particularly remarkable development, given that unstressed vowels tend to be less salient from a perceptual standpoint, and this effect may be especially magnified at the edges of words where sounds from neighboring words can interfere with clear perception of word boundaries. Indeed, in a development that has similar conditions and effect but is surely unrelated to the Greek and Slavic matter at hand for reasons of chronology, unstressed initial vowels of Latin loanwords into Albanian were lost, e.g., *mik* 'friend' from Latin *amicus*, *gusht* 'August' from Latin *augustus*, *mbret* 'king' from Latin *imperator*, etc. Putting all this together, then, on the face of it, this Greek-Slavic parallel (but no broader a scope) is, perhaps, a plausible regional Balkanism, but not at all a compelling one.

5.4.3.4 Preservation of Latin *u* as *u* (Rmn, Alb) [KS125]

In the context of a discussion of convergent features found in both Albanian and Romanian, to the exclusion of the other languages, features which, as noted in §2.1, have caused scholars to think in terms of a period of common development for the two languages since Thunmann 1774, we can note that Sandfeld mentions the fact that Latin words with *u* preserve the *u* in Albanian and in Romanian. In the former this must represent a development with loanwords, e.g., *shkurtë* 'short' (from a presumed Latin **excurtus*), whereas in Romanian, the occurrence of *u* is naturally in inherited items, e.g., *scurt* from the same **excurtus*. Sandfeld says nothing more about this convergence (but neither have other scholars since him, it seems). The interpretation that Sandfeld hints at (p. 124) for the convergence – essentially a substratum account that would treat Albanoid speakers as the ones whose shifting to Latin led to Romanian[68] – is not unreasonable, and indeed other features appear to link the speakers of these languages at a prehistoric period. Still, it seems methodologically unjustified to draw so grandiose a conclusion from a failure to change, especially since Albanian preserves PIE **u*, as in *gjumë* 'sleep,' from **supno-* (cf. AGrk ὕπνος). Even though change is inevitable and the potential for change is always present, generally change is taken to be the marked state that needs to be accounted for. All things being equal, a lack of change is the default, whether one is dealing with possible adaptation of loanwords or internal alterations of inherited forms. It seems, therefore, that this feature is not one that demands an explanation in terms of some form of contact between the groups of speakers involved and could just as easily have emerged independently in each language.

68 This is essentially the view of Hamp 1989a, 1994a, 1999; see Chapter 1 but also §5.2 for some discussion.

5.4.3.5 Treatment of Final Vowels (Rmn, Alb) [KS127, citing Meyer-Lübke 1914]

The developments to be included here are the preservation of *-e#*, the weakening of *-a#* to *-ə#*, and the dropping of *i/u/o*. They are thus highly particular, especially if they are all seen as instantiations of the same change, and moreover they involve real changes (as opposed to preservation of the status quo seen in §5.4.3.4 above). As such, even though the weakening and dropping are in general rather common changes that could arise in any language independently, this set of changes would seem to transcend the less compelling cases of the sort seen above and so might truly be of Balkanological interest. These could thus represent (true) substratum features, though limited in scope to just the substratum that fed into Romanian (and Albanian). The treatment of final /a/ in Judezmo seems to be a separate phenomenon (Quintana Rodríguez 2006: 365).

5.4.3.6 Initial *e* > *je* (Rmn, Slavic; Megl, Grk, BalkanTrk) [KS146, HS131 (citing Petrovici 1957), PA34, IS]

This feature is discussed briefly above, in §5.2, as a possible case of a localized phonological feature that cuts across language divisions, and given the nature of the effect, seems to be one that easily could carry over from one language into another, in the mouths of the same speakers. Romanian and Bulgarian are most often cited in this context, though see below concerning Balkan Turkish and Greek. The development of an on-glide with **e* (yielding *je* – **o* to *vo* also occurs, but with less regularity) is presumed to have been a Common Slavic development connected with a general on-glide prothesis with initial vowels. While it is found in West South Slavic (Slovene and BCMS) and occurs dialectally in Macedonian and Bulgarian, it is no longer present in most of East South Slavic (although the Macedonian third singular present of 'be,' spelled <*e*>, is usually pronounced [je]). But prothetic *j* before *e* is clearly a Common Slavic feature and its loss in Bulgarian and Macedonian is a later development, one which affected Turkish loanwords as well. Importantly also, from the Balkan standpoint, this prothesis occurs in Romanian, so that for Petrovici 1957, and others before him (e.g., Popović 1960: 206; Rosetti 1964b: 88; and DuNay 1977: 89), this is a Slavic feature that has worked its way into Romanian, presumably through localized bilingualism involving Slavic prior to de-jotation in East South Slavic. Thus the most likely scenario is reverse interference into Romanian via knowledge and use of Slavic by Romanian speakers. This feature is found in loanwords in Romanian, e.g., *ieftin* 'cheap,' most directly from East South Slavic,[69] but also in indigenous forms, e.g., *el* 'he' (pronounced [jel], generally taken as deriving from Latin *illum*

69 Bulgarian is the (presumed) proximate source here, with Greek as the more distant source (Middle Greek εὐθηνός, with #ε-, from earlier εὐθηνός 'thriving' (with a semantic shift perhaps via use in a sarcastic register)).

'that.ACC'). Petrucci 1999: 49–53, while dubious about others of Petrovici's claims, speaks approvingly of this Romanian feature as due to Slavic influence.

Interestingly, this feature occurs in Balkan Turkic too. Sawicka (1997: 25, see also Asenova 2002: 34 and Özkan 1996: 10) notes examples such as Rhodopian (Balkan) Turkish *jel* 'hand' (versus *el* elsewhere). This may be part of a more general tendency discussed by Pokrovskaja 1964: 73 for Gagauz of prothesis before high vowels (*ı*, *i*, *u*, *ü*), and indeed, Sawicka cites Rodopi *jis* 'trace' (elsewhere: *iz*) and Gagauz *jilik* 'first', *jüç* 'three' (elsewhere: *ilik*, *üç*). At the time Pokrovskaja wrote, however, this tendency was in retreat.

Moreover, *j*-prothesis before word-initial *#e-* is attested in some northern dialects of Greek as well. Newton 1972: 29, drawing on Phavis 1951, notes the occurrence of *je* from earlier *#e* in Chalkidiki (in northern Greece), citing forms such as [jékama] 'I did' (for expected έκαμα, a widespread regional dialectal form for this verb (StGrk έκανα, present tense κάνω, from earlier κάμνω)) and [jéxu] 'I have' (cf. StGrk έχω). Although for Phavis the basis for the development is the "stronger stress usually claimed for northern dialects," it would seem that Slavic influence cannot be discounted; the geographical restriction of the change within Greek is certainly suggestive and the north of Greece is known to have once been extensively Slavic-speaking. That is, in this case, a substrate effect as Slavic speakers shifted to Greek may be suspected as the mechanism for spread of this feature.[70] As noted above, therefore, at least as far as Romanian and some Balkan Turkic are concerned, this is a compelling local Balkanism.[71] This is quite possibly true too for Greek, though in that case it would have to be a very old influence since (virtually) all of southern East South Slavic dejotated initial /e/.[72]

5.4.3.7 *ea* or *ä* > *e* (vel sim.) when Followed by Front Vowel (Blg; Rmn; Alb; Grk) [FM, KS171, HS, PA, UH]

This feature has generated a fair bit of discussion in secondary literature, despite appearing in only a few handbook treatments. The treatments generally involve

70 Newton observes that Phavis also mentions "a pronunciation [wó] for [ó] in Kozani and other parts of Macedonia." This too is in keeping with the hypothesis of Slavic influence, since development of an on-glide with [o] is also Common Slavic. See Petrucci 1999: 52–53 for discussion of prothetic [w] before [o] in Balkan Romance, especially Meglenoromanian, but also Moldovan Romanian; he is undecided as to whether Slavic is responsible in those cases.

71 We can also note here the complex distribution of prothesis and aphaeresis in Romani dialects (on which see Boretzky and Igla 2004 and Matras 2005), including loss of Common Romani *j* in *jekh* 'one' (cf. Skt *eka-*) in Vlax and Sinti dialects, where *ek[h]* occurs. As for Balkan Turkish, as spoken in Bulgaria at least, an intriguing passage occurs in Konstantinov 1895, in the chapter entitled 'Bai Ganyo Does Elections,' where the author has a rural Turkish character say: "*Ne me ljazəm bana dajak emè*" 'Why on earth should I take a beating?', lit., 'what Q need me.DAT stick to.eat?'. The form *emè* 'to eat' represents a dative infinitive (StTrk *yemeğe*). Assuming the author correctly characterized the speech of a rural Turk in late nineteenth-century eastern Bulgaria, the absence of the initial *y* probably reflects a Turkish adoption of the East South Slavic loss of prothetic and original initial *y* before front vowels.

72 De-jotation of initial *je* > *e* also occurs in some dialects of WRT in North Macedonia, e.g., *yemiş* 'fruit' > *emiş*, which latter occurs in much of Balkan Slavic.

some apparent misunderstanding, and at the very least mislabeling is evident. The basic facts for Balkan Slavic are these: the reflex of Common Slavic *$\bar{e}$ realized in attested Slavic as *ě* (*jat'*)[73] is [e] west of a line that meanders from west of Nikopol, on the Danube, Pazardžik on the River Marica, and Goce Delčev (formerly Nevrokop) in Pirin Macedonia, down to Thessaloniki; east of that line (and also in Boboshtica-Drenova in Albania) the result is ['a] or [(')ä], i.e., a low vowel preceded by palatalization (or jotation) of the preceding consonant in some or all environments (see Stojkov 1968; Mazon 1936; Vidoeski 1962/63 for details; cf. also Gołąb 1960/61). For our purposes, the facts of the Balkan and Moesian Bulgarian dialects (northern and central eastern Bulgaria) suffice. The reflex [a] occurs only under stress and only when not followed by the sequence consonant + front vowel. The reflex [e] occurs elsewhere in the Balkan dialects, whereas in the Moesian dialects, [e] occurs in unstressed position and [ä] occurs under stress when followed by the sequence consonant + front vowel, e.g., *cv'át* 'color, flower,' *cvéten* (Moesian *cväten*) 'colored, flowered,' *cvetové* 'colors, flowers.'[74] In an apparently parallel fashion, in Romanian, one finds alternations between nonfront and front vowels (e.g., *î*/*i* and *a*/*e*) or between a backed diphthong and a front vowel (e.g., *ea*/*e*), conditioned by the presence or absence of a front vowel in the following syllable, e.g., *cuvînt* / *cuvinte* 'word' (SG /PL), *iarnă* / *ierni* 'winter' (SG/PL), *seară* / *seri* 'evening' (SG/PL). And, in Albanian there are a few similar-looking alternations, though not always obviously involving a front vowel in the next syllable, e.g., *jam* 'I am' / *jemi* 'we are,' but also *i vogël* / *e vegjël* 'little' (M/F), presumably a rather ancient umlaut process.

Schaller 1975: 84, in writing about the Romanian alternations, claims that "*Diese vokalischen Veränderungen werden also 'Umlaut,' 'Harmonisierung' oder 'Brechung' bezeichnet*" ('These vocalic alterations will thus be designated "umlaut," "harmony" or "breaking"'). His mention of *Harmonisierung* seems to be the source of two somewhat curious references elsewhere in the literature. In particular, Campbell et al. 1986: 559, in listing features of the Balkan sprachbund so as to compare the Balkans with their claim that Meso-America forms an equally recognizable sprachbund, mention "VOWEL HARMONY (or umlaut)" and say that "this trait's history is clear in Rumanian, Bulgarian, and Greek, where a stressed vowel has been influenced by the stressless vowel of the following syllable," and Thomason 2001: 108, apparently picking up on Campbell et al. 1986, lists among "the main areal phonological features" of the Balkans "some kind of vowel harmony in stressed syllables," giving its distribution among Balkan languages

73 Jat' is the name of a letter in the Cyrillic and Glagolitic alphabets that is reconstructed as having represented a low front vowel ([æ]) from a variety of Proto-Indo-European sources, including *$\bar{e}$, **oi*, and **ai*.

74 In this respect, the Balkan dialects represent the basis of Standard Bulgarian, and the Strandža dialects of southeastern Bulgaria are like the Moesian dialects. A few Rhodopian dialects retain [ä] everywhere, and most Rhodopian dialects and the Rupski dialects (Svilengrad-Haskovo) have [ä] and [a], respectively, in all stressed positions and [e] when unstressed. Northern Thrace is like Strandža and Moesia, but with closed [ę] where those latter have [ä]. See Stojkov 1968: 132–134 for additional details.

as Romanian, Bulgarian, Albanian, and Greek. It is easy to see how one might refer to these developments as a *Harmonisierung* or a "kind of vowel harmony," since there appears to be a featural matching, but this characterization and the accompanying claims about its distribution are misleading at best and simply inaccurate at worst.

First, it is not at all clear what developments in Greek are being referred to here – no examples are given and there do not seem to be any obvious cases of harmonic assimilation of vowels across syllables in Greek. While there are a few sporadic instances that look like they could involve harmonizing, such as regional dialectal όξω 'outside' and οχτρός 'enemy' from earlier (and still present in the standard language) έξω and εχθρός, respectively, or αργαλειός 'loom' from ἐργαλεῖον 'tool,'[75] it is safe to say that virtually all of the major vowel changes in the latter history of Greek are not dependent at all on vowels in adjacent syllables, at least not in a harmonic way.[76]

Second, even though one might use the term "vowel harmony" in principle to refer to a "harmonic" adjustment in which a vowel in one syllable is affected by the vowel in an adjacent syllable, so that they both come to share certain features, there are other perhaps better terms for such developments and in any case the term "vowel harmony" has a very specific meaning. In particular, a term such as "assimilation" would probably suffice for the Balkan vowel developments under consideration, at least under one interpretation of them (see below), whereas referring to these developments as "vowel harmony" would seem to be overstating the case. That is, the label "vowel harmony" generally is attached to vowel systems in which assimilatory effects involving certain vowel features range over several syllables or even over whole words, especially as complex words are built up out of constituent morphemes (e.g., by adding affixes). Turkish, Hungarian, and Finnish are among the best-known such systems, and it is clear that whatever characterization is appropriate for the Balkan facts here, they do not in any way approach the scope of those in Turkic or Uralic.[77]

Third, in Romanian there is an *o* ~ *oa* alternation where a vowel in the following syllable matters, e.g., *folositór* 'useful.MASC.SG / *folositoáre* (F.SG), also found in SDBR. This makes the Romanian (or Balkan Romance) *e* ~ *ea* look even less "harmonic."

75 It is possible, though, that the -ρ- is responsible for the vowel change of ε ([ɛ]) to α ([a]) here; note the lowering in σίδερος 'iron' ([síðɛros]) from earlier σίδηρος ([síðiros], AGrk [sídēros]).

76 Newton 1972: 30–31 does posit a change in which the height of a vowel is affected by the height of an adjacent vowel, but the effect is dissimilation (e.g., *eo* > *io*), and thus hardly "harmonic" in any sense.

77 It is also worth observing that in West Rumelian Turkish, the harmonic system so familiar from standard Turkish is not realized to its fullest extent. In some instances, this state of affairs represents an archaism, with the modern Standard system being an innovation (see Johanson 2001 and §3.2.2.9). In other instances, however, WRT has innovatively altered aspects of the harmony system under influence of contact with (i.e., bilingualism in) one or more of the nonharmonizing languages in the Balkans (see Friedman 1982c and Dombrowski 2013: 122–148 for details; also Jašar-Nasteva 1970; Ahmed 2005). One way in which vowel harmony specifically of the Turkish type finds a realization in other Balkan languages is via the borrowing of Turkish suffixes in their harmonic forms attached to particular stems – see §5.6 below for some examples and discussion, and also Friedman & Joseph 2024 on this issue of Balkan "vowel harmony" as well as vowel harmony in contact situations more generally.

Finally, when one takes the historical sources of some of these alternations into account, there is even less reason to think in terms of "harmony." In particular, the Bulgarian alternation is not a change of *ja* to *e* historically under the influence of a front vowel in a following syllable. Rather, as noted above, these reflect different conditioned outcomes of the same earlier vowel, **ě*, which is generally assumed to have been [æ] phonetically, thus neither *(')a* nor *e*. While the preservation of the frontness of *æ* before a syllable with a front vowel is an assimilatory and thus somewhat harmonic effect, the backing to *(')a* in the absence of a front vowel need not be assimilatory but could simply be a default development. In any case, it is not obviously a fronting triggered by a front vowel as the modern *(')a*/*e* alternation might seem to indicate. Describing the phenomenon in terms of a change of *ja* to *e*, as, e.g., Schaller 1975: 64 does, might be an appropriate account synchronically for at least some dialects of Bulgarian, but if this feature is to have Balkanological significance, it is not the synchronic processes involved that are relevant. Moreover, the Romanian alternation also occurs in much of SDBR (cf. Gołąb 1984a: 35; Atanasov 2002: 198), which was in contact with Balkan Slavic dialects where the outcome of jat' is uniformly /e/. This makes the (east) Bulgarian–Romanian connection look even less plausible.

As Campbell et al. and Thomason have both quite astutely and accurately noted, the conditions for these various changes are not identical among the languages that show them. Nor is it obvious that these phenomena have to be reduced to a single sort of phonological effect. Moreover, there is no reason to think that these developments have arisen due to language contact: assimilatory processes are by far the most common type of sound change cross-linguistically, so these Balkan facts lose significance with regard to the criterion of naturalness discussed repeatedly above.

Thus, even if the distribution of these changes and their characterization were more accurately portrayed, this set of developments would not be a compelling candidate as a true sprachbund feature.[78]

5.4.3.8 VV Sequences (Grk, Mac) [IS]

Sawicka 1997: 26–30, apparently alone among Balkanists, draws attention to an interesting fact about Macedonian phonology that is unique (or nearly so) within Slavic and also matches a situation found in Greek (see also Sawicka 2000a). In particular, she notes that Macedonian, and especially western Macedonian, allows vowel sequences, especially for identical vowels, to a degree unlike any other Slavic language, with verb forms such as *izbrojaa* 'they counted,' *čitaat* 'they are reading' (disyllabic [a.a] possible), *zmii* 'snakes' (with disyllabic [i.i] being usual),

78 Miklosich 1862 mentions *ea* > *e* as a feature in Romanian perhaps to be attributed to an indigenous element (*alteinheimische Element*), but he is certainly less convinced by this feature than by others; he explicitly qualifies it with the phrasing "We place lesser importance on . . . " (*Weniger Gewicht legen wir auf* . . .), suggesting that he finds it less than compelling (though intriguing enough to warrant mention).

and so on, though contraction to a long vowel does occur, especially in the southeast (see above §5.4.1.1, ii). The earliest layer of OCS had VV sequences which subsequently underwent contractions, e.g., *ii* > *i*, *ae* > *a* (Lunt 2001: 67; §4.301), so early East South Slavic did not have vowel sequences, and the Macedonian situation is an innovation. Macedonian also has three-vowel sequences in *naii* '[Ottoman] township.PL' as well as in 3PL imperfects of vocalic stems, e.g., *piea* 'they were drinking.' These all result from the historical loss of /x/ (on which, see §5.4.4.6). Moreover, in this way, Macedonian diverges from Bulgarian, so whatever development(s) reintroduced tolerable vowel sequences into Macedonian post-date the separation of Macedonian and Bulgarian and thus can be dated to (roughly) the Ottoman period. In Greek, sequences of vowels, especially like vowels, have been possible since earliest times (cf. Ancient Greek ἀάατος 'not to be hurt' (with α-privative), ἀθρόος 'in crowds; immense') though contraction was the norm in some dialects (cf. Attic ἄθρους, for ἀθρόος, with orthographic <ου> representing monophthongal [ū], the contraction product of /o-o/). In Modern Greek, there are words, generally of learnèd origin, such as ποιητής ([piitís]) 'poet,' that have a separately pronounced two-vowel sequence (here, [i.i]), and there is even a three-vowel sequence of [i.i.i], where each is pronounced distinctly, in the word αντι-ι-ικός ([andi-i-ikós]) 'anti-viral.'[79] An example of three unlike vowels in a sequence is ακουστικά βαρηκοΐας 'hearing aid(s),' with [o.i.a] in the latter part.

Admittedly, vowel sequences are not unusual cross-linguistically, especially, perhaps, when morpheme boundaries are involved (note, e.g., English *tea-y* [ti.i] 'tea-like'). Nonetheless, languages often "do something" about such vocalic clusters, and it is noteworthy in this regard that some contraction or an introduced hiatus consonant (*j* or *v*) is possible in Macedonian in some words and endings. Overall, then, on naturalness grounds, this feature may be suspect as a Balkanism.

Still, the geography within Slavic of tolerance for vowel sequences, and the availability of a source language for the presence of this feature in Macedonian, namely Greek – a source language moreover that was culturally dominant in the region, especially for Christians – at the time when Macedonian was developing its tolerance for vowel sequences, together make the parallel here between Greek and Macedonian hard to ignore completely. Thus, this feature is a plausible, if as yet unproven, localized Balkan convergence.

5.4.3.9 Vowel "Reduction" (Blg, Mac, Grk, SDBR, Rmi, Jud) [FM, KS171, EB, JF, PA, IS, UH]

This feature has drawn considerable attention, as it was mentioned first by Miklosich (though with some caution, as he uses the same wording – see footnote 78 above – to

79 This word derives from ιός 'poison, virus,' with the 'anti-' prefix αντι- and the adjectival suffix -ικό-, hence αντιικός. We owe this example to our late friend and colleague Kostas Kazazis (University of Chicago), who would be pleased, we know, to see it mentioned in a Balkan linguistics book.

characterize this development as for others he was apparently less than convinced about), and then by several other scholars, and has even been discussed herein twice already, above in §5.4.1.3 and §5.4.1.5. In those earlier sections, it is suggested that even though vowel reduction in a general sense has been talked about by some as if a pan-Balkan feature, a more realistic stance would be to focus just on the vowel raising phenomenon and to approach it from the perspective of a localistically determined case of the sort described above in §5.2. In particular, mid-vowel raising was an active process with a relatively circumscribed geographic scope in the Balkans, so that a scenario for its spread in which bilingual speakers imposed their habits of pronunciation on a second language is a conceivable one for this feature. Moreover, certain aspects of the raising meet the tests of geography and chronology for Balkanological significance.[80] Accordingly it is taken up again here, treated as a plausible regional feature within the Balkans.[81]

In particular, the earlier discussion makes it clear that the core for this change is the change (raising) of unstressed *o* > *u*, as found in Greek, eastern Bulgarian, eastern Macedonian, and SDBR. As noted in §5.4.1.5, footnote 47, we exclude Albanian from consideration here as the phenomenon is not very widely attested at all in any Albanian dialect. It is thus found in a contiguous area which forms a crescent around an area with no raising consisting of western Bulgarian on the inside and western Macedonian, southern Greek, and the Black Sea on the outside. The parallel change of *e* > *i* does not have the same extent of realization as the change of *o* to *u*, being of more limited range in Macedonian, for instance, and having a decidedly rural and provincial character in Bulgarian (unlike the widespread and fully standard raising of *o* to a more closed [o] in that language), though it is quite regular across northern Greek dialects and in SDBR. The overall distribution of both raising changes, therefore, suggests that the *o* > *u* development was the starting point for the change, presumably emanating from a locus in northern Greece most likely among speakers of the local Greek or SDBR dialects.[82] Spreading from there to speakers of Balkan Slavic, the change was generalized to take in the front mid-vowel *e* in Greek and SDBR, with a more restricted extension in Macedonian and Bulgarian. Some support for this interpretation of the original locus of the change comes from the observation of Sawicka 1997: 18, who notes that in the Macedonian dialect of Cegan, from the Voden (Grk Édessa) region of Greece, "the complete change of the unstressed /e/ and /o/ to [i] and [u] is observed." That is, in keeping with the view espoused here

80 But not all – see §5.4.1.5 for some ways in which Judezmo vowel raising is different from other Balkan realizations (and see footnote 45 above regarding Judezmo and high vowel loss).

81 We note that Thomason 2001: 108, essentially following Miklosich, cites it as among the "less widespread phonological Balkanisms."

82 It is tempting to locate this change as originating in Greek, viewing this mid-vowel raising as an extension or generalization of the earlier reduction (actually, elimination) of original high vowels (see §5.4.1.5) that characterize the northern dialects of Greek (with "raising" of *i*/*u* resulting in their deletion, i.e., raising "off the vowel space" altogether). Koneski & Vidoeski 1983: 74 write that vowel reduction in eastern Macedonian dates only from the thirteenth century, when it first appears in documents.

of local determination of the phonology due to co-territoriality and familiarity with local languages, Macedonian dialects most in contact with northern Greek dialects show most fully the Greek generalization of back mid-vowel raising to encompass front mid-vowels as well. The views of Koneski 1981: 29 are supportive, as he writes that the raising of *e*/*o* to *i*/*u* in southeastern Macedonian, northern Greek, Aromanian and Bulgarian dialects is undoubtedly the result of mutual influence among the Balkan languages. Later, Koneski & Vidoeski 1983: 74 specify eastern Bulgarian and add Romanian, also mentioning East Slavic vowel reduction as typologically relevant, in that there is a strong stress accent there that is mobile. Importantly too, in his 1981 work, Koneski extends the scope of the phenomenon to cover as well the *a*-to-*ə* development in southeastern Macedonian and eastern Bulgarian, treating that as part of the same vowel reduction process.

Under the interpretation offered here, therefore, this feature can be taken to be a true regional Balkanism, encompassing the central Balkan territory. The question of whether Judezmo vowel reduction also fits here is extremely complex, and so we only note here that such reduction does occur in Judezmo, and debates as to whether or not it is a pre-Balkan tendency (cf. Portuguese) are beyond our scope. See Quintana Rodríguez 2006: 30–69 and sources cited therein for a full discussion. On the other hand, similar vowel reductions in Romani are generally found in contact with co-territorial dialects/languages, especially in Bulgaria (cf. Boretzky & Igla 2004: II:112), and such convergence is also valid for Judezmo in Bulgaria (Grünke et al. 2023).

5.4.3.10 Aromanian Monophthongization

The Frasheriote Aromanian of the Ohrid-Struga region shows a monophthongization of diphthongs that might relate to language contact. In particular, in Frasheriote Aromanian in the Ohrid-Struga region, original /ea, oa/ have monophthongized to /e, o/, apparently due to contact with Macedonian, which has only monophthongs. For example, corresponding to Kruševo Aromanian *neádze* 'goes' is Frasheriote *nedzi*, to Kruševo *poáte* 'is.able' is Frasheriote *póti*, and to Kruševo *válea* 'river.DEF' is Frasheriote *vále* (Markovikj 2007: 22–27). The fact that Gołąb 1984a observed this tendency as sporadic among the younger generation of Kruševo Aromanian speakers when he was doing the fieldwork informing his study is an indicator that language contact was also at work in Kruševo at the time he did his research there.

5.4.3.11 Romani Centralization

One final regional development involving vowels is reported for Kalderash Romani under the influence of Romanian, as discussed and analyzed by Boretzky 1991. With specific reference to the subdialects of the Markuleš and Bunkuleš clans, Boretzky reports on not only the entry of the centralized vowels [ə] (mid) and [î]

(high) from Romanian via loanwords – vowels which are new to the borrowing speakers' otherwise "Balkan-style" five-vowel system of [i e a o u] (but see §§5.4.1 and 5.5.1.1 regarding Example 5.9) – but also the spread into native Romani words of a centralization process of *e* > *ə* and *i* > *î*, under the same conditions as found in Romanian. This is especially true for the latter change, e.g., after strong (retroflex) *r* [ɽ] and after a sibilant. Boretzky (p. 12) argues that this might "be interpreted as a transfer of on-going sound change from Roumanian onto Romani, or as a delayed process in Romani triggered by the analysis of Roumanian sound distribution" and suggests that "Romani speakers 'became aware' of the fact that after Roumanian strong *ɽ* it was normally *î*, not *i*, that occured [*sic*], and they imitated this pronunciation in their own language." In our terms, this localized contact effect would thus involve reverse interference (see §§5.1 and 3.2.1.3) from Romanian onto Romani.

5.4.4 Regional Features: Consonants[83]

As with the vocalically based regional features, the local Balkanisms concerning consonants that have been proposed in the literature are listed here, each one in its own subsection, with some pertinent discussion. See the comment at the end of §5.4.3 regarding how to interpret the listing of languages in the section heading for a given feature.

5.4.4.1 NT > ND (Grk, Aro, Alb) [FM, KS102-3, EA 109, HS131, EB51, JF47-8, PA, IS]

The basic observation here is that in a geographically connected grouping of the Balkan languages – most notably Albanian, Aromanian, and Greek, though others may be relevant (especially Meglenoromanian, where the phenomenon is sporadic, Atanasov 2002: 185) – there are developments involving clusters of a nasal consonant plus a homorganic stop (abbreviated here ND) that appear to be quite similar, both as to process and as to result, and thus are potentially interesting from a Balkanological standpoint. The finer details, however, need some elucidation, and that further discussion bears on the evaluation of these facts as a Balkanism. (See also §5.4.5.1 on homorganic nasal + stop in Macedonian dialects in Greece.)

The attention given to these developments and the precision as to their exact formulation differ in the various accounts. Miklosich 1862: 6–8 only talks about Albanian and Balkan Romance, leaving Greek out of the picture, and notes only that these languages show "*Anlaut*" ('initial sound') nasals.[84] Presumably, what he meant by this is "syllable-initial"; his two examples for Balkan Romance (in his

83 For the same reasons as with some of the regional vowel developments (see footnote 62 above), we give only this footnote mention of the reference in Sandfeld 1930: 146 to the occurrence of *fj*/*pj* > *flj*/*plj* in both Istro-Romanian and Croatian.

84 Although Romanian does not show the relevant nasal-plus-stop cluster developments, Miklosich's use of *rumunische* was the equivalent of *Balkan Romance* in the modern sense, i.e., Miklosich's *rumunische* included Aromanian.

terms *rumunische*, in modern terms, Aromanian), *îndirept* from Latin *dirigo* 'to straighten' (cf. past participle *directus* 'straight') and *împârat* from Latin *imperator* 'emperor,' involve following stops and one may presume he felt the syllabification was [î.nd . . . / î.mp . . .], although other interpretations of the data are perhaps possible, and there is dialectal variation within Aromanian (Saramandu & Nevaci 2014a).[85] His Albanian examples, *ndrekj* 'straight' and *mbret* 'king,' show a nasal Anlaut quite clearly, and the monosyllabic nature of these forms, though not insisted upon by Miklosich, is indubitable.[86]

Sandfeld 1930: 102–103 adds to this syllable-structure fact the observation, perhaps drawing in part on Skok,[87] that "*p, t et k deviennent sonores après une consonne nasale*" ('*p*, *t* and *k* become voiced after a nasal consonant'), further specifying that within Balkan Romance this development is found in some Aromanian dialects, but not in Romanian (and is sporadic in Meglenoromanian). He illustrates the development of ND from earlier nasal-plus-voiceless stop clusters (NT) with examples such as these from Greek, Albanian, and Aromanian (5.11):[88]

(5.11) a. Greek: πέντε ([pende]) 'five' (earlier πέντε ([pente]))
άγκυρα ([aŋgira]) 'anchor' (earlier ἄγκῡρα ([aŋkyra]))
τον πατέρα ([tom batera]) 'the father.ACC' (earlier τὸν πατέρα ([. . . np . . .]))

b. Albanian: *këndoj* 'sing' (from Latin *cantō*)
ngarkoj 'burden' (from presumed Latin **incar(r)icare*; cf. *carruca* 'travelling carriage' and note also *shkarkoj* 'unload' from presumed Latin **excar(r)icare*)
ngujoj 'shut and secure' (cf. Rmn *încuia* 'lock up')
ngathem 'benumb' (cf. *shkathem* 'become energetic')

c. Aromanian: *fândână* 'well, shaft' (cf. Rmn *fântână* 'idem')
mblătescu 'I weave' (cf. Rmn *împletescu* 'idem')
mângare 'eating' (cf. Rmn *mîncare* 'idem')

Most other scholars – except for Sawicka 1997: 49–59, 2014: 68–75 (see below) who devotes more space to nasal-plus-stop clusters than all other treatments combined and extends the scope of the discussion considerably – give ND only

85 Miklosich cites the forms in Cyrillic, which was still official for Romanian at the time. In Romanian, Cyrillic ѧ normally denoted < î > + nasal, but to avoid any possible ambiguity, Miklosich gives a form with ѧ + Cyrillic <n, m> in parentheses. Note too that *îndirept* might not be from an initial *d-* of Latin but rather from a presumed prefixed verb **in-dirigo*, since from *directus*, Romanian has simply *drept*. There are, however, clearer examples, e.g., *ndeṣ*, etc. 'thick [of hair]' (Saramandu & Nevaci 2014a: Map12; cf. also Map 93; Rmn *des*, Megl *des*, Lat *densus*, cf. Cioranescu 1958–1966: s.v.).

86 Testimony from native speaker consultants generally confirms this; see also §2.2.3, footnote 8.

87 Sandfeld cites Skok 1928, who has a brief comment (p. 410) on Aromanian: "*Wir wissen auch dass gewisse arom. Dialekte auch die Sonorisierung der nach n folgenden Konsonanten gleich dem Neugriechisch kennen. Diese Erscheinung hat sich bekanntlich auch auf das Alban. ausgedehnt*" ('We know too that certain Aro dialects know the voicing of consonants following *n* like Modern Greek. This phenomenon has evidently been extended in Alb').

88 For Aromanian, this type of voicing is found basically in Epirus and the southern part of Aegean Macedonia (Saramandu & Nevaci 2014a: Map 93).

brief mention, and based on the facts as laid out above, one is thus inclined to consider both the syllable-onset fact and the voicing of the stop to be possible areal features. Indeed, Friedman 2006a states the convergence as follows: "Greek, Albanian, and Aromanian all have initial nasal + homorganic stop and a tendency to voice stops after nasals." As becomes clear below, this judicious statement is probably the closest to what can be said definitively about the Balkanistic status of nasal-plus-stop clusters, though some further clarification is needed. As is so often the case in the Balkans, the chronology of the developments leading to these two main facts about syllable structure with ND and about post-nasal voicing is crucial to their assessment.

Regarding the chronology of the emergence of ND clusters in the Balkans, the change of NT to ND occurred in Albanian no earlier than post-Roman times, to judge by the fact that early Latin loans into Albanian are affected in the same way as inherited items; e.g., *këndoj* 'sing' in (5.11b) is just like inherited *ndej* 'stretch out' from Proto-Indo-European **en-ten-* (cf. AGrk τείνω 'stretch,' < **ten-jō*). As for Aromanian, the fact that it diverges from Romanian on this feature (note (5.11c) above) points to an emergence of this feature in Aromanian after it split off from the rest of Balkan Romance, a date usually assigned to about 1000CE.[89] Moreover, the fact that the development is not represented consistently in all Aromanian dialects (nor in Meglenoromanian), and, moreover, the change is most consistently represented in the Aromanian south, indicates that it developed after dialectal differentiation within Aromanian itself as well as the separation of Meglenoromanian. In Greek, the postnasal voicing of original NT clusters[90] could well have been an extension of the tendency toward postnasal voicing of stops in Greek of the Hellenistic and Roman periods (Bubenik 1989),[91] and there are spellings in papyri from the seventh century CE that indicate the occurrence of post-nasal voicing.[92]

89 Rosetti 1973: 169 dates the separation of Romanian and Aromanian to the tenth to twelfth centuries (rejecting Philippide's 1923–27: 2.225–30 dating of the separation to the sixth century).

90 Note that NT clusters did occur in Ancient Greek; Sawicka 1997: 49 – a fine work in general that we have obviously benefited from enormously throughout this chapter – offers an account of Ancient Greek stops that is so inaccurate that one wonders if it was garbled at some point in the transmission from her main source (Tarabout 1985) to her text or even from her manuscript to the printed page. For the record, then, she misstates things when she refers to "an interesting distributional phenomenon" in *Ancient* Greek whereby "voiced stops /b/, /d/, /g/, /dz/ [*sic*; this last sound never occurs in Ancient Greek/VAF-BDJ] occurred always only after nasal sonants whereas voiceless stops – without such a prenasalization." Similarly, contrary to the facts of Ancient Greek, she suggests therefore that voiced stops could not occur independently, e.g., in initial position. She continues erroneously to say that "intervocalic voiced stops underwent fricativization and only in the word-initial position and in some clusters [did] they remain occlusive"; here it is true that Ancient Greek voiced stops became fricatives but that happened in all positions except after nasals and thus also in initial position. This last fact, coupled with the post-nasal voicing of NT to ND, led to the situation in later Greek (and into the twentieth-century standard language) wherein voiced stops were restricted to post-nasal contexts (which is most likely what Sawicka was attempting to describe). Sawicka 2014: 69 eliminates some of these errors.

91 This tendency is seen in Pamphylian, an Ancient Greek dialect attested in Asia Minor during the Hellenistic period, for it has <δ> for earlier **nt*, e.g., 3PL present ending <-οδι> from earlier **-onti*.

92 Tonnet 1993: 45–46 points to seventh-century spellings such as <πεμμπι> for Ancient Greek <πέμπει> (['pempej]) 'sends,' where the first <μ> stands for a nasal, so that the remaining letters <μπ> must stand for something else. However, they cannot be a voiceless [p], for that would

We can safely surmise that as far as Greek was concerned, it surely was completed by the beginnings of the Middle Greek period (c. tenth century CE).

These facts suggest late Hellenistic to early Middle Greek as the temporal locus for the Greek developments but a somewhat later date for the Aromanian ones. It is hard to place the Albanian developments into a time frame more precise than simply post-Roman contact, but we note that "post-Roman" here can mean really any time between roughly the third century CE and the thirteenth century.[93] This, combined with the dialectological evidence, means that it is very likely that the Aromanian ND developments were due to contact with Greek. Such is the position that Saramandu 1984: 433 takes, for instance – in noting that in the Pindus dialects of Aromanian (in Greece), the opposition between voiceless and voiced stop is neutralized in that there are groups of stop plus nasal only with voiced stops, not with voiceless stops (thus, e.g., *mb*/*nd* but not **mp*/*nt*), he says this is the result of "*influenţa exercitată de neogreacă asupra aromânei*" ('the influence exerted by Modern Greek on Aromanian'). It should be added, though, that depending on when and where Albanian fits into the overall picture of ND developments, contact with Albanian could have played a role for Aromanian as well, or, perhaps, vice-versa, with Aromanian affecting Albanian. Given the dialectal distribution seen in Saramandu & Nevaci 2014a, however, this seems unlikely, except, perhaps, in connection with Çam expansion into Epirus. Still, one cannot rule out a greater role for Aromanian in some aspects of the later development of Albanian, just as, in the view of Gołąb 1984a, it played a much greater role than is usually supposed in the development of Macedonian. Moreover, the nature of contact between Greeks and Albanians in the post-Roman and early Byzantine period does not seem to have been as intense or as much in the direction of the sort of bilingualism that might affect pronunciation as was Balkan contact during the Ottoman period or later (see below). Thus it may well be that the Greek developments here are to be separated from the Albanian ones, even if the Aromanian and Albanian facts on the one hand, and the Aromanian and the Greek facts on the other, can be seen as connected.

It is relevant to note here that Greek is unlike Albanian and Aromanian in the matter of nasal Anlaut to words and syllables. In the passage from Ancient Greek to Modern Greek, earlier clusters of NT and ND both converged on ND;[94] with the loss of (most) unstressed initial vowels (see §5.4.3.3), some ND clusters were thrust into word-initial position (e.g., AGrk ἐντρέπομαι 'I feel compunction' > ντρέπομαι)

certainly be spelled simply <π>, so that the digraph <μπ>, he argues, likely represents a voiced stop [b] here.

93 There is no relevant evidence from early Slavic loanwords to delimit the upper bound on "post-Roman" in this context (assuming sixth to seventh centuries for the entry of Slavs into the Balkans), since all instances of -VNT- in pre-Slavic changed (usually to a nasalized vowel, -*ṼT*-) away from the very context in which post-nasal voicing might be found in Albanian. The form *opangë/opingë* 'leather sandal,' a loanword reflecting Slv **орьнькъ* (cf. Serbian *opanak*, Mac *opinok*) with a fleeting vowel that reduced to give a new *-nk-* cluster), need not indicate a late date for NT > ND, but probably reflects on-going effects in the aftermath of the earlier sound change (see footnote 98 below concerning phonotactics vs. "permanently active" changes).

94 This is the most prominent position where Ancient Greek voiced stops did not fricativize (see footnote 90 above).

while others (as in (5.11)) were word- or phrase-medial (e.g., AGrk πέντε 'five' (with [nt]) > MGrk πέντε, with [(n)d]). Although the development of these ND clusters across Greek dialects is rather complicated, the "classic" statement of the distribution of their pronunciation for the standard language and for many of the dialects (see, e.g., Mirambel 1933: 157; Householder et al. 1964: 20; Newton 1972: 96) is that in initial position the nasality is lost, leaving a pure oral stop, while in medial position the nasality is retained (thus, standard Modern Greek ντρέπομαι ([drépome]) but πέντε ([pénde])). It is likely that passage of #εντ- ([#ent-]) to #ντ- ([#d-]) occasioned a period with a word-initial sequence starting with a nasal and continuing into a stop, i.e., *#nd-*. Whether that was a nonsyllabic nasal, as in Albanian and Aromanian, or realized as prenasalization on the stop (i.e., [#^{n}d]) is hard to say, and in any case, the ultimate widespread loss of the nasality in Greek is quite unlike Albanian and Aromanian. This might then constitute another reason, other than chronology, for treating the Greek situation as distinct from that of these other languages taken together (even if individually there might be a connection).[95]

Some additional facts about Greek word-initial voiced stops, as well as medial ones, need to be made clear. There are pronunciations both in dialects and in the standard language whereby some degree of nasality is to be heard with initial voiced stops. Arvaniti & Joseph 2000: 144–145, 150, for instance, note (and verify instrumentally) light prenasalization on a few tokens of voiced stops in their sampling of spoken data from standard language speakers, e.g., μπροστά ([^{m}brosta]) 'forward.' It is not at all certain, though, that these represent the maintenance of a trace of the earlier nasal consonant, for they could just as well represent a spelling pronunciation on the part of the speakers (given that voiced stops are spelled with a digraph of nasal and voiceless stop, e.g., <μπ->) or phonetic facultative nasality emerging from a slightly different timing of the onset of voicing and oral release of the stop (not unlike colloquial American English [^{m}baj] for 'Bye!'). Thus even though Sawicka 1997: 52, following Setatos 1969: 44, cites the possibility of "unmotivated nasals . . . in strongly emotionally marked utterances" (e.g., imperatives such as [ndisu] (ντύσου) 'GET DRESSED!!!' versus more usual [disu] (ντύσου) 'get dressed!'), these cannot be taken as evidence for nonsyllabic nasal Anlauts being maintained from Middle into Modern Greek;[96] emphasis is exactly the sort of context where nasal "enhancement" to a stop might be expected and could well be phonetically determined. Moreover, there are dialects that show some nasality initially. For instance, Sobolev 2009a notes for the Peloponnesos the Slavic color term *belo* for 'white ram or lamb' (Map 56) and the Albanian term *bartsa* for 'white-bellied goat' (Map 88) with light nasality on what was an initial simple voiced stop in the source languages. Similarly, the Aromanian *Mbãliot* 'pre-Frashiote inhabitant of Bela (di Suprã [Mac Gorna Belica])' sometimes realized as

95 Here it is worth noting that preservation of initial nasality is considered characteristic of Greek in the Giannitsa region (Papadamou & Papanastasiou 2013), where contact with Aromanian and Albanian (as well as Macedonian, cf. §5.4.5.1) was especially intense.

96 The verb in question here comes from AGrk ἐνδύνω 'put on, don.'

Măbăliot, represents a prenasalization that has been attributed to Greek influence. In some instances, the nasality is a retention, and one can suspect that contact with a co-territorial language may play a role, as, for instance, with the robust nasals in ND clusters in the Greek of southern Albania, even in word-initial position (e.g., ό, τι μπορώ 'what I-can' pronounced as [óti mboró], where Albanian influence may be manifest; cf. C. Brown & Joseph 2012). The retention as a feature of northern Greek dialects, which are or were also in contact with Aromanian and Albanian (Papadamou 2019a; Papadamou & Papanastassiou 2013) could also be significant.

With medial ND in standard Modern Greek now – and historically in some regional dialects too (e.g., Cappadocian, cf. Dawkins 1916: 69, 81, and Aegina, cf. Thumb 1891: 107–108) – there is variation between nasal (ND) and nonnasal (D) realizations. In contemporary usage among speakers of standard (Athenian) Greek, Arvaniti & Joseph 2000 found that the nonnasal realization is nearly categorical now for younger speakers. Still, variation has been noticed for some time in the literature (Newton 1972: 97 and Kazazis 1976 are some early mentions of it), and, to judge from the evidence of early twentieth-century recordings of rebetika songs (Arvaniti & Joseph 2004) and regional dialect differences, it has historically been a part of Greek phonology for some time. Nonetheless, there must have been a period in Greek where the occurrence of a pure voiced stop intervocalically was phonotactically highly marked if not outright impossible, but whether that period was long enough to allow Greek to have an effect on, say, Aromanian is not clear though not really counter-indicated either.

Variation that is reported by Sawicka 1997: 54 for Aromanian word-initial #ND may be important here. Papahagi 1905, 1974 has cases of #NT- in Aromanian (thus referring to dialects different from the southern ones which Saramandu 1984, cited above, mentions), such as *mpartu* 'separate' (from Latin *impartire*), *nklid* 'close' (from Latin *includere*), and Sawicka herself mentions doublets such as [dilikat]/[ndilikat] 'delicate,' noting further that Gołąb 1984a describes for "the dialect of Kruševo a reduced vowel . . . before such an initial cluster, cf. *ᵊnklidu*, *ᵊmpartu*." See also Saramandu & Nevaci 2014a: Map 93. The dialectally variable appearance of a vowel and a nasal is reminiscent of what the Middle Greek situation must have been like before the widespread resolution to initial voiced stops without a nasal prop.[97]

Albanian too shows variation, though between ND and N outcomes, with forms such as *mret,* etc. being attested for historically prior *mbret*. It is Tosk (and the Debar/Dibra dialects) that preserve nasal + stop clusters, while Geg and the transitional dialects lose them (Gjinari 2007: Map 31). Thus, it is precisely in the region most heavily influenced by Greek (or, in the case of Debar/Dibra, Slavic)

97 Regarding synchronically variable forms at odds with historical outcomes with ND, we discount the relevance for the Balkan ND situation of initial clusters of *m* with other consonants reported by Sawicka 1997: 55 for Romanian, in which a reduced form [m] of the weak dative personal pronoun *îmi* gives onsets such as [mdaj] 'you give me' (for *îmi dai*) or [mspune] 'he tells me' (for *îmi spune*). The lack of homorganicity and the variability suggest a recent development that thus is not pertinent to the historical Balkan situation. See below too for variable aspects of ND in Albanian.

that nasal + stop is most stable. Conversely too, and also characteristically Tosk, historically canonical *#mr-* clusters can develop an epenthetic stop, as in *mrekullueshëm* / *mbrekullueshëm* 'miraculous,' from Latin *miraculu-*. While an epenthetic stop developing in such a cluster is not in itself particularly unusual (see §5.4.4.7 for discussion), it does mean that there are actually two sources of variation with realizations of ND clusters in Albanian, and they are correlated with proximity to Greek.

It is relevant to the phonotactic interpretation of the resulting ND situation that that later loans are affected in various of the languages. For instance, Turkish *kantar* 'balance' => Albanian *kandar*, Italian *conte* 'count' => Greek κόντες ([ko(n)des]) 'count,' and English *computer* => Greek κομπιούτερ ([ko(m)bjuter]) can be taken to indicate that phonotactic restrictions on where plain voiced stops could occur remained in effect in the aftermath of the sound changes in question.[98] Note too that some loans with medial plain voiced stops were borrowed into Greek with nasality; Newton 1972: 122 reports pronunciations such as [somba] 'stove' for the "Turkish" loanword *soba* (see BER VII: s.v. for the etymological complexities). Since the carry-over of phonotactic restrictions, or conversely, reverse interference involving phonotactics, in second-language learning situations can occur, these developments are plausible contact traits, assuming the chronology was right for, e.g., Greek influence on Aromanian, and then Albanian–Aromanian interaction as well.[99]

In this regard, facts concerning Arvanitika in Greece and Romani in Albania are quite pertinent. Although they reflect presumably somewhat more recent developments, they show the power of phonotactics in situations involving bilingualism.

Sasse 1991: 61–62 reports for Arvanitika a shift with ND clusters due to the influence of Greek, which (in his formulation) has no pure voiced stops but only voiced stops accompanied by some nasality; as he puts it, the Arvanitika voiced stops "*tendieren . . . zur Pränasalierung*" (p. 61, 'tend . . . towards prenasalization'). At the same time, he says, in the traditional clusters of N + D, "*wird . . . das nasale Element geschwächt*" (p. 61, 'the nasal element . . . is weakened'). This leads to a neutralization of the original distinction between, e.g., /#b-/ and /#mb-/, both being realized as [#^{m}b-], and, he says, this neutralization occurs occasionally

98 The reference here to phonotactics is somewhat equivalent to the claim of Çabej 1988: 392 that, as Sawicka 1997: 52 puts it, "sonorization of /t/ and /k/ after N took place [in Albanian] very early in the prehistorical period, and it has been operating permanently up to now." Still, there are two exceptional contexts that neither approach accounts for satisfactorily: the 3PL imperfect ending *-onte* resists voicing, as does the *-k* of the admirative after a nasal ending, the abbreviated participial form it attaches to in forming the admirative (thus *qenka* 'it [to my surprise] is!') has [ŋk] not [ŋg]).

99 There are occasional instances of [mb] for etymological [b] in Judezmo. Crews 1935: 234 cites the verb *asumbir* 'go up,' noting that the regular form is *asu~~b~~ir*; etymologically the medial consonant here is a voiced stop [b] and spirantization of earlier voiced stops intervocalically is quite regular. Judezmo does however preserve /b/ as a stop in initial position (Crews, p. 179–180), so that one might wonder whether, were a stop pronunciation of medial /b/ somehow maintained in this word, Greek-like phonotactics could lead to the [mb]-variant *asumbir*. As this is an isolated instance, however, Crews's explanation (p. 234) that it is due to analogical influence from another verb, *sombaer* 'seduce, tempt,' is probably preferable.

among even the oldest speakers. Tsitsipis 1998: 25, footnote 3, gives examples of a different development, but one also consistent with Greek influence, in which ND clusters in Arvanitika move in the direction of the Greek D resolution noted above with the loss of the nasal, and he says this is mainly something that younger speakers do, exactly the age-group that in Greek is losing the nasal. Sasse too notes differential age-related variation, in that for younger speakers he observed, there is variation between pure nonnasal realizations and light prenasalization word-internally for canonical pure voiced stops (e.g., /robə/ "*Morgenmantel"* ('robe') coming out as [robə] or [ro^{m}bə]) and for words with canonical N + D clusters there is variation between full nasal and pre-nasalization (e.g., /məngəl/ "*Ärmel"* ('sleeve') => [məŋgəl] / [mə^{ŋ}gəl]); he generalizes further (p. 62) concerning variability in the speech community overall:

> Die Tendenz zur Neutralisierung des Kontrastes Verschlusslaut: Nasalverbindung ist nicht bei allen Sprechern gleich stark ausgeprägt. Einige Sprecher weisen sie sehr früh auf, andere behalten die Opposition bis zum totalen Sprachwechsel bei. Im allgemein findet sich die Neutralisierung eher bei Männern als bei Frauen.
>
> The tendency towards neutralization of the contrast stop : nasal cluster is not similarly strongly realized in all speakers. Some speakers show this very early, others maintain the opposition up through the complete change of language. In general, the neutralization is found rather in men than in women.

Hamp 1989b: 201, too, remarked on variation between ND and D in the Arvanitika of a speaker from Liópesi that he observed in the 1950s; this speaker showed the effects of "the modern intense Hellenization . . . and binds up the phonetic voiced stop with nasality" (with this account based in part on Hamp's analysis and observation about Greek of the time as having voiced stops only after a nasal).

And Sawicka 1997: 56, bringing a new language into the arena regarding ND, refers to Balkan Romani developments. In particular, she reports that "Balkan Romani dialects living in the Albanian Diaspora [show] unmotivated nasals . . . sometimes, cf. [žamba] 'frog,' [andresa]" 'address,' adding that "the clusters in question can occur in the initial position of a word, cf. *ngarav* 'I carry.'" Since variation with ND in Albanian is between ND and D, as discussed above, it is not clear that Albanian is the source of these variations. In the case of *ngarav*, the form comes from the loss of an initial vowel, the forms *angarav* and *ingarav* both occur commonly in the Balkans (along with many variants; Boretzky & Igla 1994: s.v.), and loss of initial *a-* is a Romani dialectal feature. In the case of *žamba* and *andresa* the possibility of a Greek intermediary exists. The form *žamba*, a borrowing from Slavic *žaba*, is found as far away as Caló, and could represent a very old Slavic borrowing via a Greek intermediary, and *andresa* would be more recent. Sawicka's mention (p. 59) of ND/D variation for Albanian, though, might suggest that these Romani forms reflect phonological patterns of Albanian, as the dominant language, spilling over into Romani (reverse interference).

The situation is a bit more complicated in Slavic, which Sawicka also brings into the spectrum of Balkan ND developments. She mentions but quite rightly dismisses as not overly significant, loans from local non-Slavic languages found in "Serbian and Macedonian dialects existing in the Albanian and/or Greek diaspora" (1997: 56), such as *fambrika* or *juŋguslavija*, but she feels that for *mr* > *mbr*, "the process itself is borrowed [e.g.] in Macedonian villages of Southern Albania" that have [mbleko] 'milk' (vs. [mleko] elsewhere). More recently, Sawicka & Cychnerska 2018: 73 note that these examples are old, and that the current dialects no longer show these features. Still, given the nature of phonetic tendencies, an argument for contact-induced change is not straightforward. And, in her cases like [amberika] in "the Slavic-Muslim [i.e., Goran – VAF/BDJ] village of Brod" (Sawicka 1997: 57) for [amerika] 'America,' we suspect that hypercorrective pressures are responsible (note that the canonical form is with [m] here so it cannot be a phonetically derived instance).

With regard to the development of secondary nasal vowels – and then full nasal consonants from them – in Macedonian dialects in Greece and Albania where some originally nonnasal vowels became nasalized, based on inherited Slavic nasal vowels before consonants that develop into VNC sequences (see §5.4.1.1, iv above and §5.4.5.1 below), analogical extension (akin to hypercorrection) can also be a cause. Sawicka 1997: 57–58, discussing cases like [mangła] 'fog' in such dialects, from **mъgla* (cf. [magła] in most of Macedonian), argues for the relevance of syllable structure ("it divides clusters in which a sonant stands between an obstruent and a juncture"). However, given that there would have been fluctuation between Ṽ and VN, with the quality of the vowels being similar, i.e., centralized, that fluctuation could easily have been extended to other centralized vowels, a sort of phonetic analogy (in the sense of Vennemann 1972).

For Sawicka 1997: 49–59, then, other aspects of ND developments are seen as convergent across the various languages, and ND/D variation is a relevant part of the overall Balkan picture. She mentions the following specific points (p. 59):

- Unmotivated N before T are regular in Southern Greek, they are frequent in Albanian dialects and happen in Arumanian
- Unmotivated stops after T [*sic*; N is surely intended here] often occur in Albanian dialects, especially in the clusters [mr], [ml] – > [mbr], [mbl]
- Tendency towards functional equivalence of ND and D occurs in Greek. Double forms testifying such a tendency are also met in Albanian and Arumanian.

These parallels, however, can be taken to reflect highly local effects induced in bilingual environments, in keeping with Friedman's (2008a) formulation concerning *Balkan phonologies*. In particular, some of these parallels surely have different sources. The "unmotivated N before T" cases, such as Albanian *fambrika* 'factory' for *fabrika* (with Albanian as a likely source for *fambrika* in dialectal Balkan Slavic, Sawicka 1997: 56) or Aromanian *fambricã* 'factory' (Papahagi 1974: s.v.) could reflect local hypercorrections in the face of other cases of ND/D variation. The epenthetic stop effect (unmotivated stops after nasals) is likely to be

phonetically induced in each instance, given that this is a phonetically very natural process involving the timing of featural release in transitions from one sound into the other (cf. Hock 1991: 118) and as such might be expected to be found in language after language independently. Still there is still a case to be made for the possibility of shared elements of phonotactics, cf., e.g., Aromanian *mbăliot* '[non-Frasheriote] inhabitant of Bela di Suprã (Mac Gorna Belica),' which shows an older habit of prenasalization, while the current local name of the village, *Belã*, does not (cf. also Sawicka 1997: 56, 2014: 43, 59; Sawicka & Cynerska 2018: 41, 47; Sawicka & Dargiel 2018: 36). Overall, then, while the individual cases may be reactions to situations caused by language contact, it is hard to say that the appearance or not of a nasal or a stop is itself a direct outcome of contact.

In terms of how to evaluate this development, therefore, it is important to note first that the NT > ND change in itself (not in any further ramifications), to the extent it is shared among these different languages, must be considered with some degree of caution based on the criterion of naturalness; that is, from a cross-linguistic standpoint, this is a common change, and, being assimilatory in nature, it is not at all unexpected on phonetic grounds. Indeed, the examples that Sandfeld 1930: 103 and Sawicka 1997: 55 cite from Italian dialects (e.g., *tembo* for *tempo* 'time') and other languages (Sandfeld notes Oscan *ander* from Indo-European **entero*- (cf. Latin *inter*), for instance[100]) certainly suggest that a contact-based explanation for post-nasal voicing need not be posited for any of the languages in the Balkans as a whole, although some of the dialectological distributions are suggestive. Similarly, some of the ways in which variation arises involving ND, especially the appearance of innovative epenthetic stops in clusters of, e.g., -*NR*-, to give e.g., -*NDR*- versus -*NR*-, involve phonetically well-motivated processes that recur in many languages independently. For that matter, too, the reduction of ND to D would seem to have phonetic plausibility on its own, without needing recourse to contact. It seems therefore that we might be dealing here – especially insofar as the Albanian developments are concerned, and especially given the chronological considerations discussed above and the likelihood of Greek influence on Aromanian – with a change that originally in Albanian was independent of anything seemingly similar in Greek or Aromanian, but, in Tosk, was then reinforced, or nudged in a particular direction, perhaps at the very least simply gaining stability, through contact with Aromanian and Greek.

The further convergent aspect of the ND situation, namely the occurrence of the nasal syllable-initially, even if not fully evident in current Greek and dialectologically limited in Albanian and Aromanian, is nonetheless a striking fact that is at odds with notions of preferred syllable structure cross-linguistically.[101] As such, it

100 Sandfeld also mentions Thracian, which raises the possibility of a substratum effect with ND clusters in the Balkans. Sawicka 1997: 50 briefly mentions this, citing approvingly Di Giovine's 1980 refutation of such a view; among other things, the chronology of ND in Greek and Aromanian speaks against any involvement of a substratum.

101 For instance, the Sonority Sequencing Principle of Clements 1990 posits that syllables universally are structured with increasing sonority, so that a syllable that starts with a (more sonorous) nasal that is followed by a (less sonorous) stop, as in Albanian *mbret*, would have a dispreferred structure.

clearly passes the test of the naturalness criterion and is thus a solid candidate for being a significant contact-induced feature. Moreover, the area in question here is a fairly compact band of contiguous languages/dialects in regions where the effects of bilingualism are evident.[102] Finally, the facts of more recent contact-induced effects from Greek onto Arvanitika and, perhaps, Albanian onto Romani show that developments involving ND, especially of a general phonotactic nature (i.e., whether voiced stops can stand alone or need to be preceded by a nasal), can be passed on via bilingualism. Putting all of that together, the nasal *Anlaut* aspect of ND in the Balkans would seem to be a good candidate for a local phonological Balkanism in the sense identified by Friedman 2008a. By extension, therefore, even with concerns that the naturalness criterion might raise, the post-nasal voicing may be considered a reasonable candidate as well.

All in all, then, the developments with nasal-plus-stop clusters reported by Miklosich and Sandfeld and endorsed by Friedman stand as Balkanologically significant, though perhaps not as thoroughly interrelated as might at first seem.

Moreover, given that three different branches of the Balkan languages are represented in these developments – admittedly only partly so in the case of Balkan Romance (southern Aromanian but not the rest of Balkan Romance), surviving only in Tosk in Albanian, and only somewhat more weakly present in the case of Greek – and given further that this number is as many languages as are represented with the postposed definite article, for instance, among morphosyntactic features, it would not be unreasonable to view this Balkanism as worthy of more general attention. While such a move might run the danger of being incorporated into the "scorecard" approach to Balkan convergences that we find generally misleading, it could also serve the useful function of giving the list-makers more reason to think in terms of a more nuanced approach to Balkan phonological convergences.[103]

5.4.4.2 Elimination/Creation of Dental or Palatal Affricates (Grk; Rmi; Arv; Mac; Aro, Jud; Trk) [Skok 1928: 410, cited in KS 103; PA, IS]

Within Balkan Romance, it is noteworthy that Aromanian has dental voiceless and voiced affricates *ts dz* where Romanian has palatal affricates, so that Aro *tsints* 'five,' *fudzire* 'flight' compares with Rmn *cinci* ([činč]), *fugire* ([fudžire]). Sandfeld, citing Skok, interprets these correspondences as showing "*le passage en aroumain de č et dž en ts et dz*" ('the passage in Aromanian of *č* and *dž* into *ts* and *dz*'), and, again apparently basing himself on Skok, suggests that influence from

102 Note in particular the discussion in §5.3 above regarding *d*/*ð* in Greek loanwords in different Aromanian dialects.

103 Reflecting on the practice of Balkan scholarship, we cannot help but wonder if the absence of Balkan Slavic from the more widespread ND developments led to less discussion in the literature of nasal-plus-stop clusters than, for instance, the stressed schwa (which, as noted in §5.4.1.6, has little chance of being a significant contact-induced feature). After all, many, perhaps most, practicing Balkanists over the years have come into Balkan linguistics from a Slavistic background. See now, however, Sawicka 2014 and Friedman 2018a.

Greek might be responsible, since Greek, as noted in §5.4.2.1 above in general (in the standard language at least) has only dental affricates and regularly substitutes them for the palatal affricates of donor languages in loanwords. On the face of it, this development seems like a plausible candidate for a phonological convergence between Aromanian and Greek. Moreover, to the extent that Greek dental affricates (see below) derive via a "depalatalization" from earlier palatals – this is the generative interpretation of Greek dialect differences given by Newton 1972: 128ff. to model dialects with palatal affricates compared to those with dentals – parts of Greek may have participated in a shift like that posited by Sandfeld for Aromanian.

The correspondence between Aromanian and Romanian is certainly real regarding these affricates, and, moreover, there are cases of languages in the Balkans in contact with Greek where affricates are realized in the Greek manner, i.e., as dentals.[104] Igla 1996: 190, for instance, documents this phenomenon for Romani in Greece in the dialect of Agía Varvára, a suburb of Athens, in which *ts*/*dz* are found for common Romani palatal affricates.[105] The same development holds for many speakers of Arvanitika (Sasse 1991: 58–59).[106] Moreover, the mechanism of change involved would be a now-familiar one (see §5.2 above) involving reverse interference from speakers' second but dominant language, in this case Greek, onto their ethnic heritage language. Similarly, in some of the Bulgarian dialects from Greek Thrace, one finds *št* > *st*, *žd* > *zd*, and *č* > *c* (Bojadžiev 1991: 90). For Macedonian, among the younger generation of speakers in Greek Macedonia, there is a tendency to replace palatals with corresponding dentals under the influence of Greek (VAF field notes). In the younger generation of Macedonian speakers, the change may be a symptom of Greek becoming dominant. See also footnote 56 above on Balkan II (Burgudži) Romani dialects, whose most salient feature is the replacement of palatals (and jotated velars) with dentals.

Nonetheless, there is reason to be skeptical regarding the nature and cause of the Aromanian affricates, and in particular to question whether foreign influence of the sort envisioned by Sandfeld and Skok is really at work here. Thus the status of this

104 Sala 1971: 128, note 96, citing various earlier sources (Subak 1906: 153; Wagner 1930: 17; and Crews 1935: 216), notes that some Judezmo dialects, even ones in contact with Greek, e.g., in Thessaloniki and Izmir, have [dʒ] replacing [dz]. This is surprising, as it is opposite to the dentalization of palatals that might be expected given contact with Greek. Still, the directionality here is not clear, since these sounds represent palatalized *d*, and much depends on the early outcome of palatalized *d* in Old Spanish, i.e., whether it was [dz] or [dʒ]. Opinions seem to differ. If it is a matter of [dz] replacing an older [dʒ], then Judezmo is like the cases cited here. But if it was indeed the reverse, with [dʒ] replacing [dz], then hypercorrective pressures may have been at work, in the context of speakers showing variation between [dʒ] and [dz], whatever the historical directionality may have been for this change.

105 The etymology of *čamčali* 'eyelash' from Georgian *c'amc'ami* 'idem' (Friedman 1988a), probably involves a change of *c* to *č* in Romani itself. See also footnote 56.

106 Recall too from §5.4.2.1 that the dental affricate "pattern" of Modern Greek seems to have affected some of those Romani dialects in sustained contact with Greek in their adaptation of loanwords with affricates as well. Here it is noteworthy that the Romani dialect of Ágios Athanásios, now part of Sérres in Greek Macedonia, lacks dental affricates completely (Sechidou 2011: 12).

development as representing a Balkan convergence at a localized level is somewhat uncertain.

First, in particular, Sandfeld himself says that *č* is preserved in Aromanian in other contexts, or, perhaps better, when arising from a different source: "*il faut pourtant observer que l'aroumain a gardé intact č dans les combinaisons čo et ču sorties de tio et tiu*" ('one must yet observe that Aromanian has kept *č* intact in the combinations *čo* and *ču* which came from *tio* and *tiu*'). That is, *ti* gives *č*, which then remains before a back vowel. If Greek influence as conceived of by Sandfeld and Skok were responsible for the occurrence of *ts* in Aromanian, one would not expect a differentiation between *či* in the word for 'five' and *čo-* from earlier **tio-*.

Second, while it is indeed the case that Greek has only dental τσ ([t^{s}]) and τζ ([d^{z}]), this is true really only for the standard language (and the dialects on which it is based), as suggested above. In fact, palatal affricates are found in various regional dialects, in some from earlier velars and in others from earlier dentals. While in the northern zone, in the general area where Aromanian speakers are found in Greece, velars are mostly intact, though somewhat fronted, before front vowels, there are some northern dialects in which velars do develop into palatal affricates in this environment, especially around Aráchova (near Delphi, Newton 1972: 131), and the affricate outcome may be more widespread.[107] Moreover, in other parts of the northern dialect zone, e.g., Siátista (Macedonian Sjatišta), in Greek Macedonia, a predominantly Greek-speaking town with an Aromanian presence, palatal affricates have developed out of earlier dental stops before front vowels (e.g., *jačí* 'why' from earlier (and standard) γιατί ([jatí]); see Newton 1972: 145). While this last fact is interesting in the light of Sandfeld's comment about Aromanian *čo*/*ču* from earlier *tio*/*tiu* and raises the thought of external influence (in one direction or the other), the details of the affricatization of dentals in Aromanian suggest otherwise, since the processes in the two languages are actually quite different. In Aromanian, stem-final *t* and *d* (so also *k* and *g*) in fronting environments, e.g., before the *-i* of noun plurals or the *-i* of the second-person singular present tense, turn into dental affricates, and the triggering vowel disappears (coalescing with the stop in the formation of the affricate); thus the palatal outcome that Sandfeld remarks on is found only before back vowels; in Greek, by contrast, this process seems to affect only *-t-* and it yields a palatal outcome with the vowel intact if stressed or ultimately deleted if an original unstressed *i*, but the vowel loss there is part of the regular and widespread northern Greek loss of unstressed high

107 Palatal affricates from velars occur also in the northern Aegean island of Lesbos, within the northern dialect zone (Newton 1972: 131), and in the Greek of southern Albania (C. Brown & Joseph 2012, 2013), presumably independent developments from that found in Aráchova, given the geographic distance between the areas. Thumb 1912: §17 includes in the areas showing "palatalising of a κ before *e* and *i* (*y*)" the mainland areas of Locris and Aetolia, in the northern dialect zone; however, he does not distinguish tsitakismos (*k* > dental affricate *ts*) from softening (palatalization) as Newton does, counting both *ts* and *č* as "palatalising." Therefore Thumb's wider recognition of palatalization in the north may be a matter of applying different criteria. And, as Peter Trudgill (p.c.) has reminded us, one cannot necessarily trust all dialect descriptions regarding palatalization; one has to wonder if Greek linguists, most of whom were speakers of the standard language with only dental affricates, could always hear a palatal accurately.

vowels. There is thus little parallelism in detail between the processes that led to the Greek and the Aromanian dental affricatizations. Moreover, it is not necessarily the case that the passage from velars to dental affricates must even go through the same sort of stage(s) of palatalization as the passage of velars to palatal affricates. For instance, at the point of the second Slavic palatalization, in which velars before new front vowels (namely **ā̆* (jat') and **ī* from earlier **oi*) developed eventually into dental affricates, the palatal affricates which had previously come about as a result of the first Slavic palatalization from earlier velars before original front vowels all stayed as palatal;[108] it is generally assumed that what kept the outcomes distinct was the development of a distinction between strident palatals (before original front vowels) and mellow palatals (before secondary front vowels), with the mellow ones ultimately dentalizing. Thus while one could take Newton's model of Greek dialect differences seriously as a historical account, so that Greek and Aromanian would have both undergone a depalatalization of the same type of palatal affricate, this is not a necessary step for Greek.

This same reasoning holds for Aromanian. In fact, there is no reason to think that Aromanian must be the depalatalizing innovator in the difference between it and Romanian regarding the affricates. In words like *tsints*/*cinci* 'five' or *fudzire*/*fugire* 'flight,' with affricates from Proto-Romance velars before front vowels, it could well be that *ts*/*dz* represent the older outcome, or, perhaps more plausibly, that Proto-Balkan Romance could have had fronted *k'*/*g'* from which Aromanian and Romanian innovated their respective affricates each on their own. And, finally, if contact with speakers of some other language is involved in some way in the Aromanian developments here, it could be that the other language was Slavic, since, as just noted, the early Slavic palatalizations could well have provided the relevant influence.

There are Judezmo dialects in which *dz* was simplified to *z* (Sala 1971: 128), and these tend to be Eastern dialects, for instance Istanbul Judezmo. Quintana Rodríguez 2006: 367–375 gives detailed maps of the fate of voiced dental affricates in Judezmo. The isoglosses for de-affrication form a bundle that begins south of Epirus, bulges northward to include Bitola and Skopje, goes south of Sofia, and then more or less runs along the Danube. There is even a change of *ts* to *s* in Hebrew words such as מצה (matsā) 'unleavened bread' that follows a similar, albeit not identical, isogloss (cf. also Quintana Rodríguez 2006: 110).

Note too that new instances of *ts* are found in WRT from Slavic loans, e.g., feminine suffix *-itsa* (which sometimes occurs as *-içа*; Jašar-Nasteva 1970). Finally, to return to a variety of Greek, in Ottoman-era Edirne Greek (OEGrk), as described in Ronzevalle 1911, 1912, various loanwords from Turkish occur that have palatal affricates in Turkish, and consistently have palatals in OEGrk, e.g.,

108 Some mergers of outcomes of the first and second palatalizations in Slavic did occur in some of the languages (e.g., Polish merges the outcomes of **x*), but for the most part the two results are kept distinct and certainly were so at the time they arose.

[tšát pát] ‘imitation of a noise that a body makes while breaking up, or two bodies that clash’ (1911: 287),[109] from Turkish *çat* ‘sudden sharp noise’ or [dʒüdʒés] ‘dwarf,’ from Turkish *cüce* (1911: 286); these loans were adopted without adaptation, as was the case with other Turkish features in OEGrk (see, e.g., §4.3.3.2 on the borrowing of postpositions as postpositions). In the other direction, in the Turkish of Komotini (Trk Gümülcine) and nearby villages in Greek Thrace, Turkish palatals are pronounced as alveolar, e.g., [kats] for [katʃ] (StTrk *kaç* ‘how many’); cf. Petrou 2018.

Thus although there is much of interest regarding Balkan affricates and some reasonably secure localized Balkan convergences involving affricates (see especially §§5.4.2.1 and 5.7 for more on affricates), the particular one that Sandfeld and Skok focused on seems better explained without recourse to language contact than with it.

5.4.4.3 Presence/Absence of *ð θ γ* (Grk, Alb, Aro, Mac, Rmi, Jud) [KS 103–4, EA 108–9, HS 131 (regarding *ð* only), JF, PA, IS]

It has been noted in the literature that there is agreement among Greek, Aromanian, and Albanian, alone among the Balkan languages, in the occurrence of voiced and voiceless phonemic interdental fricatives (*ð θ*) and to some extent also, the occurrence of a voiced velar fricative (*γ*) as well. In Judezmo, the voiced interdental and velar fricatives are allophonic and were brought from the Iberian peninsula. The qualification regarding *γ* is needed since as far as Albanian is concerned, this sound occurs primarily in Arvanitika, although *γ* also occurs in some of southern Geg as the reflex of *ð* (see below, footnote 115) and some of Northeast Geg as a result of voicing of /x/ (see §5.4.4.6).

Some Macedonian dialects, too, show some of these sounds, judging from reports from the first half of the twentieth century and more recent literature as well, thus spanning several generations of speakers and probably differing social conditions regarding bilingualism: Boboshtica (Mac Boboščica, in southern Albania), according to Mazon 1936: 46 and Vidoeski 2000: 249, has *ð θ* in loanwords from Albanian and also from Greek, and has extended *ð*, in place of *d*, even into some words of Slavic origin (cf. also Steinke & Ylli 2007: 311 and Afendras 1968: 70, 109, who cites as well Šramek 1934). In the Macedonian of Albanian Gora, *ð* occurs exclusively in Albanian loanwords (Steinke & Ylli 2010: 57), and it is apparently absent from the Goran dialects in Kosovo (Mladenović 2001). As for the other Macedonian dialects of Albania, it appears that neither fricative occurs in Prespa, Vrbnik, or Golobrdo (Steinke & Ylli 2007, 2008). The fricatives occur in the Macedonian of Nestram (Grk Nestório) in Greece according to Šmiger [Schmieger] 1998: 56–58. Moreover, both Hill 1991 and Dvořák 1998,

109 In the original: “*imitation du bruit que fait un corps en se brisant, ou de deux corps que s’entrechoquent*.” This is actually the meaning of *çat* alone. In modern Turkish, as in other languages (see (4.15) and accompanying footnotes), *çat pat* means ‘a little bit’ referring to an ability to speak a language. It can also mean ‘now and then’ or ‘here and there.’

describing the Macedonian of Gorno Kalenik (Grk Áno Kalleníki) and Popəlžani (Grk Papagiánnis), respectively (villages in Greece in the Lerin (Grk Flórina) district, near the border with North Macedonia), mention the occurrence of *ð* and *γ* in these dialects, mostly, but not exclusively, in loans from Greek.[110]

A similar situation is found in Aromanian of Greece, as discussed in §5.3, and revisited below. Aromanian in Albania (spoken from Tirana southward), likewise has *ð θ γ*, while that in North Macedonia and Bulgaria does not (Neiescu 1997). Likewise some Romani dialects in Greece and Albania have *ð θ* in loanwords and those in Greece also have *γ,* which later is sometimes realized in North Macedonian as <y> (e.g., Igla 1996: 12; Cuvata 2006: s.v.; VAF field notes).

The languages involved are found in a geographically contiguous area; moreover, Romanian does not have any of the sounds in question, nor does the rest of South Slavic, so that Aromanian and dialectal Macedonian diverge from their closest linguistic relatives in this regard. Furthermore, it is clear that language contact has something to do with this phenomenon, since most or all instances of these sounds in Aromanian and Macedonian, whether in Greek or Albanian-speaking areas, are in loanwords from Albanian or Greek (see §5.3 above for some discussion), and the same is true of Arvanitika, as far as *γ* is concerned. And these sounds in dialectal Macedonian have a clear contact basis too. Thus the presence of these sounds on the face of it would seem to be a good candidate for a local Balkanism. Still, a somewhat closer examination of the facts raises some doubts as to their Balkanological significance, although ultimately the verdict is that there is something noteworthy going on here.

First, the sources of *ð θ* in Albanian, Aromanian, and Greek are all quite different, and moreover, the chronology of the appearance of these sounds differs for each language. As just noted, these sounds generally occur in Aromanian only under contact with Greek or Albanian and thus emerge quite late in the language's history (though some dialects – see §5.3 – extend these sounds into native words, e.g., *ðimtu* 'wind' from Latin *ventus* and *θeamină* 'female, feminine' [Papahagi 1974: s.v.; Neiesucu 1997: 99]; cf. Latin *femina*; cf. North Aro *feaminã* [Cuvata 2006: s.v.]). For Greek, however, *θ* and *ð* are the results of regular sound changes that predate any serious contact with other Balkan languages, deriving respectively from Ancient Greek [t^h] (orthographic <θ> in the Greek alphabet) and [d] (orthographic <δ>) via changes that took place at least in the Hellenistic period and possibly earlier.[111] As for Albanian, *θ* is the regular outcome of the Proto-Indo-European voiceless unaspirated palatal stop **ḱ* (cf. Tosk *thëri* 'louse' from **ḱonid-* (AGrk κονίς)) and *ð* results regularly from Proto-Indo-European palatal **ǵ*, e.g.,

110 Hill 1991: 24–25, for instance, cites native words like *graðo* 'the town' (*grad-* 'town' plus postposed article), with intervocalic *ð* (for etymologically expected *d*; cf. standard Macedonian *grad-ot* 'town-the').

111 Exactly when the modern fricative pronunciation of Ancient Greek voiceless aspirated <θ> arose is somewhat controversial, but a date of no later than the late Koine period is reasonable; so also for <δ> (see Bubenik 1989: 189ff. for some discussion). Moreover, there are earlier indications of movement away from the stop pronunciation of <θ> in some Ancient Greek dialects; fourth century BCE Laconian Doric, for instance, has <σ> for <θ>, suggesting a fricative pronunciation.

Tosk *dhëmb* 'tooth' from **ǵombho-* (AGrk γόμφος 'bolt'), as well as **ǵh*, and from PIE **d* and **dh* word-internally (as in *pjerdh* 'fart' from **perd-* (Skt *pard-*). These developments are pan-Albanian and thus they must be very early as far as Albanian is concerned, before the major dialect divisions arose and thus well before any period of significant contact with Greek. It is therefore hard to assume that the occurrence of these sounds somehow reflects the wave-like spread of a single innovation across what is now Greek-/Albanian-/Aromanian-speaking territory.

Second, as with the case of stressed schwa discussed above in §5.4.1.6, one has to ask if the mere occurrence of particular sounds in one language would be sufficient to "steer" speakers of other languages in contact with that language in the direction of developing that same sound in their own language. Thus, even if one wanted to argue for contact with post-Koine Greek being somehow responsible for Albanian developing these sounds, one has to wonder why, for instance, the occurrence of [θ] in Greek at that point would induce an Albanoid **k'* in the direction of [θ] and not simply [k], a sound also present in Greek and in some respects closer phonetically to *k'* than *θ* is. Presumably, then, PIE (Albanoid) **k'* was already somewhat *θ*-like, but if so, then Greek need not be invoked to explain the Albanian sound.[112]

Thus, the occurrence of *ð θ* in both Albanian and Greek most likely represents the result of completely separate and independent developments which are therefore without Balkanological significance. It may well be that the persistence of these sounds in these languages, since they seem to be in general fairly "stable" – to the extent that one can characterize anything in language in that way – and not the "target" of widespread dialect shifts or the like,[113] is due to a period of mutually reinforcing bilingual contact among speakers of these languages, but that is not the same as explaining the occurrence of the sounds in the first place as having a contact origin.[114]

As for Aromanian, however, the discussion in §5.3 above demonstrates that contact with Greek is the source of its *ð θ* and that familiarity with Greek on a localized basis is responsible for the unadapted acceptance ("adoption") of these sounds in borrowings. The same holds as well for instances of *ð θ* that entered Aromanian in Albanian loanwords, for again, familiarity with co-territorial Albanian would have made a difference. A similar scenario can be invoked for the *γ* that first entered Aromanian in Hellenophone territory via loanwords from

112 It is also unlikely that Albanoid influence was responsible for the appearance of [θ] in Greek; the chronology of the emergence of Greek [θ] speaks against that, although it is not certain when predecessors of Albanian speakers entered the Balkans. Also, though, the Greek change leading to [θ] was part of a general shift of voiceless aspirates to fricatives, since [p^h] and [k^h] shifted to [f] and [x] respectively, whereas the Albanian change is restricted just to [k'].

113 Intervocalic voiced fricatives, including *ð*, are lost in Cypriot and parts of the southeastern Greek-speaking world, making it tempting to think that being outside of the Balkans per se encouraged such a development; however, Cypriot maintains [θ] even without contact with Albanian. Note too that positing, quite reasonably, a lenited stage of [ð] as intermediary to the loss of intervocalic /d/ in various Macedonian dialects would mean that being in a contact zone with languages with [ð], as Macedonian is, did not aid in the retention of the sound.

114 See §3.2.8 for some discussion of contact-induced retention.

Greek. Moreover, as noted in §5.3, some of these sounds have spread to words in Aromanian not of Greek origin, including ones inherited from Latin or borrowed from Slavic. Again, intense contact with Greek and the familiarity with Greek that the resultant bilingualism entailed can be said to be responsible for these facts about Aromanian. Finally, the same scenario applies to the case of Arvanitika *γ*, which entered in Greek loanwords (cf. Hamp 1989b: 200) and remained unadapted due to familiarity with Greek on the part of the Arvanitika speakers, and to the Macedonian adoption of loanwords without alteration. Mazon 1936: 46 notes the unadulterated adoption in Boboshtica (Mac Boboščica) especially among younger speakers, to whom "*le dh albanais est familier*" ('the Albanian dh is familiar'); cf. also Hill cited above on *ð* in Macedonian and Clopper 2017 on the effects of familiarity. Thus to some extent contact is clearly at the heart of these central Balkan developments with *ð θ γ*, even if not in a way that would make this a three-way local Balkanism; most likely, Albanian, except for Arvanitika *γ*, is to be excluded from consideration here and the Greek-Aromanian facts are to be taken to have, or better, to have had, some Balkanological significance as a local Balkanism. The reason for qualifying even this statement is that, as Friedman 2006a has noted, among speakers of Aromanian who do not know Greek or Albanian, especially younger speakers in North Macedonia, these fricatives are often replaced by stops. Thus, for example Gołąb 1984a: 40 observes that there are "two subsystems in the Kruševo dialect: an older, cultivated and conservative, in which in which numerous Greek loanwords preserved the original Greek pronunciation [. . .] and a younger, colloquial one, in which for the Greek spirants the corresponding Romance stops are substituted." Again, though, familiarity with the donor language is a crucial element, and that can only take place on a highly localized basis.

We can also note here that some Albanian dialects in Hot and Malësia e Madhe on both sides of the Albanian–Montenegrin border either replace /ð/ with /ɬ/, or the two are in free variation (Gjinari 2007: 93). This replacement of /ð/ by /ɬ/ is unique in Albanian, and occurs in a region of intense historical Albanian–Slavic contact.[115] Also, a simplification of Albanian *ð θ* to *d t* occurs in the dialect of Mandrica in Bulgarian Thrace under Bulgarian influence (Sokolova 1983, cf. Hamp 1965).

Judezmo is also relevant here. The spirantization (or lenition) of intervocalic /d/ and /g/ to /ð/ and /γ/, respectively, had already occurred before the expulsions of the Jews (Pharies 2007: 88; Penny 2002: 76; Lloyd 1987: 327). Initial /g/ could also spirantize. Although the exact trajectory and spread of spirantizations is complex and debated, it is fair to say that spirantized /ð/ /γ/ were brought to the Balkans by the Jewish refugees whose language would become Judezmo, and that the change

115 A further development with interdental fricatives in Albanian is found in another small region, the triangle between the rivers Erzen and Shkumbî with Kus at the apex and Peqin and Elbasan as the respective west and east points of the base, where *th*/*dh* ([θ/ð]) are replaced with *x*/*γ*, respectively, thus giving a language-internal source of [γ] in Albanian. Other than these two small regions, interdental fricatives are quite stable in Albanian.

/ð/ /γ/ > /d/ /g/ occurred precisely in areas where Slavic and Romanian were dominant colloquial languages. Moreover, the isogloss for the preservation of spirantized /ð/ goes up the Black Sea Coast, where Greek remained significant into the Ottoman period (cf. §1.2.3.3). See Quintana Rodríguez 2006: 88–93, 113–114, 128, 377–379 for details on the various distributions, which are complex. Still, the overall generalization is striking.

5.4.4.4 *mj* > *mnj* [mɲ] (Grk, Arv)[116] [KS 104]

This feature is discussed above in §5.2 as a likely example of the effects of reverse interference from Greek as a second language into Arvanitika as a first(-learned) language. Even though it appears to be a phonetically quite trivial development, in that the *j* can be seen as pulling the labial nasal in the direction of a more interior articulation, thus assimilatorily moving it towards a palatal realization,[117] the geography of the convergence, i.e., involving Arvanitika and Greek, the "availability" of Greek as a source language for this development, and the social fact of Greek being the dominant majority language in Arvanitika-speaking territory together make this a good candidate for a contact-induced change. Thus even though only Sandfeld mentions it, and even though the *mnj* development occurs elsewhere in Albanian (see footnote 116), we judge this to be a very plausible local Balkanism in the phonological domain.

5.4.4.5 *sk* > *št, šč* /__[+front] (Rmn, Blg, Mac, NGrk) [FM 7, KS 146]

Sandfeld 1930 was virtually alone among Balkanists for decades in observing a parallel between Balkan Romance and East South Slavic in the development of an earlier sequence *sk* before a front vocalic element, *i*, *e*, or *j*. Amongst earlier scholars, though, Miklosich 1862: 7 makes an oblique reference to such a parallel and Seliščev 1925: 49 has a related observation (see below). More recently Hamp 1989a: 44 mentions it, and Petrucci 1999: 53–57 has some extended discussion. What Sandfeld observed was that in Romanian and Bulgarian this sequence yields *št*, as in Rmn *ştiu* 'I know,' from Latin *scio*, and Blg *štit* 'shield,' from earlier **skeit-*. Echoing Seliščev (though without citing him), he goes on to mention the outcome *šč* in southern Aromanian and co-territorial southwestern Macedonian, presumably as a further related parallel, although he does not explicitly call it that (whereas for

116 The same change occurs in the Albanian dialects of Gur-Lurë (Peshkopi), Mackull (Mat), Sohodol (Peshkopi), and Tanushaj (Dibra). These four points are contiguous. At the two edges (Mat and Dibra) *mjekër* 'chin; beard' > *mnjekër,* while in the center (Peshkopi) *mnjekër* simplifies to *ɲjekër*. See also §5.2.

117 As a near-parallel to this change, Proto-Indo-European **m* yields Greek ν ([n]) when followed by *j*, as in κοινο- 'common,' if it is from **kom-jo-* (cf. Chantraine 1977: s.v.); presumably the same change affected syllabic **m̥* before *j*, as in βαίνω 'come' from **g*^{w}*m̥-jo-*. And, Sawicka 1997: 38 notes some instances of *mj* to *mnj* in Romanian dialects, citing Caragiu-Marioteanu 1975, e.g., *miel* 'honey' => [m^{n}el] (so also Istro-Romanian, with [mɲe]; see Neiescu 1980), though this seems to be part of a broader process of labial palatalization, since it is discussed along with the realization of palatal coarticulation effects on /p/ and /b/ as well.

Seliščev, it is a matter of a direct parallel, of the sort termed here a local Balkan convergence – see above in §5.2 and footnote 7). Moreover, Sandfeld's examples of *šč* do not quite bear out his suggested parallelism, other than in the existence of the phonetic sequence *šč*, since the Aromanian *šču* 'I know' that he cites does indeed reflect an earlier *sk* (Latin *scio*), but his Macedonian examples of *ščo* 'what' and *praščam* 'I send' reflect earliest Common Slavic (and PIE) **kito-* (CoSl **čito*; cf. attested OCS *čьto*) and **pratj-* respectively. What these examples do show is that the picture is actually somewhat more complicated in that *šč*/*št* in these South Slavic varieties derive not just from **sk* in a palatalizing context but from **tj* as well as **kt* before front vowels and *j*. Thus for Slavic one might well talk in terms of a neutralization of these consonants in this context, whereas in Balkan Romance, one finds different outcomes (note in particular, Romanian <ţ> (= [t^s]) from **t* before a front vowel (as in *ţara* 'land' from Latin *terra* 'land').

Augmenting this picture, but ultimately lending it some clarity, is the fact that in northern dialects of Greek, i.e., varieties co-territorial with southwestern Macedonian and with Aromanian, *sk* before a front vowel yields *šč* (as in [ščuli] 'dog,' with secondary rounding and backing of the root vowel,[118] compared with standard Greek σκυλί). This adds another language to the group showing *sk* to *šč*, so that there is a somewhat larger area to reckon with in evaluating the validity of this potential convergent feature but still a relatively compact area of contiguous or co-territorial speech communities.

It must be borne in mind, though, that the palatalization of *k* to *č* is quite widespread around the Greek-speaking world though admittedly absent from some dialects, especially the standard language, parts of the Peloponnesos, and parts of the north (so Thumb 1912: 17; Newton 1972: 131ff.) with the further change of *s* to *š* if it happens to precede a *k* that is undergoing palatalization to *č*. This fact raises the possibility that what Sandfeld and others have drawn attention to is merely a natural phonetic development, and one can point to other parallels, such as Italian, with its development of *šš* out of Latin *sk* before a front vowel that is likely to have passed through a stage of *šč* (so Petrucci 1999: 55–56).

Nonetheless, given the occurrence of *sk* > *šč* precisely in co-territorial dialects and not elsewhere is highly suggestive of contact, especially since Aromanian and Greek *šč* are limited precisely to the regions where Macedonian shows it as well. And despite the phonetic naturalness of the developments in question, it can be argued that just because something is "natural" does not mean that it is not due to contact (as seen with *mj* > *mnj* in §5.4.4.4). Thus the extension of Sandfeld's observation to take in the *šč* outcome in some of Balkan Slavic and Balkan Romance (and also northern Greek dialects) is reasonable as a local Balkanism.

With regard to the narrower Romanian-Bulgarian *št* convergence, it is striking, and seems to be a less common outcome of the fronting of a *k*. Indeed, Petrucci

118 This vowel change is most common in Greek in the vicinity of labials and velars and sonorants; see Joseph 1979 for some discussion. Alternatively, the palatalization of *sk* here could reflect a late retention of the front round quality of υ, as in Ancient Greek (a value suggested by Newton 1972 for Greek into as late as the tenth century). See also §5.4.1.1.iii.

1999: 55–56 feels that Slavic influence on Romanian may have occurred in the last stage, guiding *šč* to its final Romanian form of *št*, and others (see especially Trummer 1983) have commented on an isogloss within the Balkans of "*št* versus non-*št*," pointing to the distinction between Šćakavian and Štokavian dialects of BCMS, for instance, as well as *št* outcomes of *sč*' in a good part of East South Slavic, and also the Romanian developments. Thus, even if not conclusively due to contact, inasmuch as the input sequences that yield Bulgarian *šč* are far broader than those for Romanian, this narrower convergence is also a plausible and quite reasonable local Balkanism.

5.4.4.6 Loss of *x*/*h* (Mac, Alb, WRT, Rmi; BCMS) [IS 34–36]

Most of the Balkan languages had a velar fricative [x] or glottal approximant [h] in their consonantal inventories at an early stage of their development. Bulgarian, eastern Macedonian, and Greek, as well as much of Albanian (and also most of BCMS), have retained this sound, and some Romani dialects have both. Indeed, presence of this sound is one of the phonological characteristics of the Balkans mentioned by Feuillet 1986 (item (5.10g) above). It is interesting, then, that the loss of these sounds might be considered noteworthy too, and there are relevant facts here suggesting contact-induced change. As noted by Sawicka 1997: 34–36 and Friedman 2006a (see also Friedman 1982c: 14), in dialects of various languages as spoken in western North Macedonia, specifically standard Macedonian and the western dialects on which it is primarily based, Albanian, and West Rumelian Turkish, *x*/*h* is generally lost or replaced, especially by *f* (and see above, §5.4.2.6, regarding the shift of *h* to *x* in Arvanitika).

For instance, in Macedonian, Common Slavic **xleb*- 'bread' gives standard Macedonian *leb*- and **xubav*- 'beautiful' gives *ubav*- (compare Bulgarian *xubav*-), **snъhá* 'daughter-in-law' gives *snaa,* but in some contexts earlier **x* develops into *f* or *v*, as in aorist forms such as *vidof* 'I saw' and *vidofte* 'you (all) saw' (orthographic standard *vidov, vidovte*), *muva* 'fly,' from earlier *vidox* / *vidoxte* / *muxa*. Standard Macedonian has re-acquired a distinctive /h/ in recent loanwords, e.g., *hotel* 'hotel' and other secondary means, e.g., Church Slavonicisms like *duh* 'spirit,' but otherwise it is absent from a large sector of the lexicon, although the distribution is complex.

In Albanian, *ftoft* 'cold' for standard *ftohtë* is typical of northern Geg (Gjinari 2007: 125). On the other hand, the realization of *shoh* 'I look' and *njoh* 'I know' as *shof*/*njof* is typical for central and southern Geg as well as the transitional dialects (Gjinari 2007: 183). A variant of *njoh* that appears in Standard Albanian is the related lexeme *njoftim* 'announcement, information.' In many dialects, however, /h/ disappears altogether or in specific positions (initial, medial, and/or final). As Sawicka 1997: 34 reports, "in a number of the Northern, Eastern, and Central Albanian dialects [x] is being replaced or lost in any position . . . for example in Devoll, Berat, [and] Dibër . . . in Pogradec, Skrapar and some Çamerian dialects [x] is lost only in word final position" (see Gjinari 2007: 89, 182–185 for details, but also Jusufi 2011: 174–175 for corrections).

Further, in West Rumelian Turkish, one finds, e.g., *amur* 'pie' and *kave* 'coffee' for *hamur*/*kahve* elsewhere, with occasional other developments, such as initial *je-* for *he-* in Gagauz, e.g., *jer* 'everyone' for *her* elsewhere in Turkic. And, Matras 2009: 230 refers to "Macedonian Turkish [where] /h/ in most positions, especially initially, undergoes weakening and frequent omission as a result of contact with Macedonian and Albanian, which lack such a phoneme [dialectally]."

Elimination of /x/ by loss or replacement extends as well into BCMS. Ivić 1991: 90–101 gives an overview of this and related phenomena in Štokavian.[119] In Bosnia and the Sandžak, preservation of /x/ is typically Muslim (Bosniac), although sometimes Catholics (Croats) also preserve it, and in Mostar, so do the Orthodox (Serbs). Elsewhere (e.g., the Banat, parts of Istria and Dalmatia), preservation of /x/ is typical either for Orthodox or Catholics. Ivić connects the preservation of /x/ in BCMS with contact with Turkish, Albanian, or Hungarian and its loss with contact with Romance. In those BCMS regions where preservation of /x/ is a Muslim ethnic marker, it could be that the perception of /x/ as Turkish was a part of the picture (cf. Blevins' 2017 concept of *perceptual magnets*). Given the heterogeneity of /x/-loss in BCMS, it would appear that different explanations apply in different places, and ethno-religious marking differentiation played a role (cf. Gjinari 1975: 97, mentioned below; see also Greenberg 1996b).

Here we can also note one other development involving the elimination of /x/. Ivić 1985: 161 observes that among the Muslim speakers of the former Serbo-Croatian in the Plav-Gusinje (Alb Guci/Gusî) and the southern Sandžak, in Montenegro, /x/ undergoes a change to *γ* in initial and medial position, e.g., *γoćeš* 'want.2SG,' *muγe* 'fly.GEN.SG.' Ivić considers this to have occurred under the influence of co-territorial (or formerly co-territorial) Albanian dialects which, he writes, have the same change.

Interestingly, Sawicka 1997: 34 notes that Gjinari 1975: 97 mentions another religiously based isogloss, namely that in Devoll (in southern Albania), and she reports it thus: "[x] is maintained only by Moslems."[120] However, on the other side of the Balkans, in Romanian, one finds not the loss per se of *x* or *h*, since Common Balkan Romance, and Proto-Romance for that matter, had no such phoneme,[121] but rather the transformation of **x* of Common Slavic into *f* in the earliest layer of loanwords from Slavic into Romanian.[122] Given that *h* is

119 Preservation of inherited /x/ is characteristic in Čakavian and Kajkavian. See especially R. Greenberg 1996a, also Friedman 2006a, for additional details on /x/ in Štokavian.

120 Gjinari actually writes about [h] being lost in the speech of Christians but the point is essentially the same (and [x] and [h] do not contrast in Albanian).

121 Latin *h* disappears in all the Romance languages, and thus is safely assumed to be absent from Proto-Romance. Moreover, there is evidence from the Classical period involving dialectal Latin (e.g., *edus* for *haedus* 'goat' cited by Roman grammarians (see Joseph & Wallace 1991b for discussion and references) and from occasional hypercorrections involving [h], even some that became standard, such as *humerus* 'shoulder,' to suggest that loss of [h] began relatively early in Latin.

122 A later layer of borrowings creates Romanian /h/ out of Common Slavic **x*; see footnotes 58 and 59 for some discussion. Note also other sources of [h] in Balkan Romance: Aromanian has *h* from

generally considered to be a rather weak sound, certainly from a perceptual standpoint, but also lacking a certain degree of articulatory energy, any loss of the sound cannot be considered too unusual a phenomenon, and certainly numerous presumably independent parallels to the loss of *h* can be cited, e.g., Cockney English, early Ionic and Aeolic Greek, Greek more generally in the Hellenistic era, and Latin (see footnote 121), among others. Thus an explanation such as that implicit in the discussion of Gjinari 1989: 77–78, 184–188, 305–306 that the developments affecting *h* in Albanian are due to structural pressures, in particular the fact that *h* was isolated in the consonant system with no voiced counterpart, may be unnecessary. Moreover, given the acoustic similarities between the velar fricative *x* and the labial fricative *f* – both have a low second formant showing "gravity," in the feature system of Jakobson et al. 1967, and citable parallels showing their interchangeability, e.g., the development of forms such as *laugh* ([læf]) in English (note OEng *hlæhhan*, Grm *lachen*), the development of *h* to *f* is not at all surprising.

It is likely therefore that the early Romanian treatment of *x* in loanwords is unrelated to other developments with *x* elsewhere in the Balkans, but that these developments elsewhere, inasmuch as they involve contiguous geographic areas, may well be significant and contact-related. Especially striking is the agreement between western Macedonian dialects and Albanian dialects in the same or adjacent areas (for which the evidence of toponymy is decisive, cf. Ylli 2000) in having *f* for *h*, and that particular development may very well be a local Balkanism, a true regionalism. To this can be added that in those dialects of Romani that have lost the *x*/*h* contrast, there is a tendency for original /h/ to be lost, while /x/ is retained as /h/. While the distribution of this simplification is complex, it is striking that it is consistently realized in the various Romani dialects of North Macedonia (Boretzky & Igla 2004: Map 23).

5.4.4.7 Epenthetic Consonants in Clusters with Sonorants (Mac; Blg; Grk; Rmi) [IS 45 PA 38]

In several Balkan languages, regional dialects show the development of an epenthetic consonant, generally a stop, in clusters involving sonorants (nasals, rhotics, laterals), in combination either with other sonorants or with sibilants. Thus, northern Greek dialects (Newton 1972), where the loss of high vowels has created some appropriate clusters with sonorants, have forms such as the proper name Απουστόλτς ([apustolts]) from earlier (as in the standard language) Αποστόλης (via [apustóls]), πκάμψου ([pkámpsu]) 'shirt' from earlier (and standard) πουκάμισο (via [pkámsu]), μπλάρ ([mblár]) 'mule' from earlier (and standard) μουλάρι (via [mlár]), among many others. Similarly, Albanian dialects have *embri* 'the name' for standard *emri*, and in Romani one finds for the instrumental

Latin *f* before front vowels (as in *her* 'iron' < *ferrum*), in loans from Slavic (as in *hrean* < older Slv *hrěn* 'horseradish'), and in onomatopoeia (e.g., *ham* for a dog's bark). In Meglenoromanian, /h/ is from the local Greek and Macedonian dialects (Atanasov 1990: 160).

plural of *manuš-* 'man' the form *manušenca* 'with the men,' with *c* from an epenthetic *-t-*, built from the oblique plural stem *manuš-en-* combined with the instrumental suffix *-sa*. Sawicka 1997: 45 calls this phenomenon a "buffer consonant" but then talks about this as "affricatization of fricatives after sonants." This label is descriptively suitable in that an affricate (or perhaps better, affricate-like sequence) results from the development of a stop before a sibilant, as in *ls* > *lts*, but there are good reasons for labeling it differently.[123] For one thing, it is conceivable that what results in such cases is a biphonematic cluster and not a true (monophonematic) affricate; the difference between the two is real but subtle,[124] and is not always detectable just from listening, even with a trained ear, and not always observed by linguists providing a transcription.[125] Moreover, some linguists (e.g., Newton 1972: 214, who is not alone in this regard) require that the stop and fricative be homorganic in order to qualify as an affricate, so that under such a view, Sawicka's own example of Greek [pkampsu] would require a different treatment. Also, as Sawicka herself seems to recognize, the sibilant "phenomenon is connected with" the [mblar] type, so that affricate-like sequences need not even be involved.

Most importantly, the emergence of such a buffer consonant is so common cross-linguistically and so readily understandable in terms of the physiology of the articulatory transition from one sonorant to the following sonorant or sibilant (see §5.4.4.1 and Hock 1991: 118) that this phenomenon has to be suspect as a contact-induced Balkan feature on the grounds of naturalness. Thus, as interesting as it is to see this consonantal epenthesis recur in language after language in the Balkans, there is no guarantee that contact is at all involved in it.

5.4.4.8 Laterals [FM 7, IS 32–33]

As (5.10c) above (§5.4.2) suggests, with its common Balkan inventory of sounds – even if not to be taken seriously as a Balkanism – laterals are among the sounds occurring in all the Balkan languages. However, although (5.10c) mentions only *l*, there are actually far more laterals to be found, and they present some intriguing convergences at both broad and local levels. Laterals were noticed as sounds of potential interest regarding possible convergent phonology as early as Miklosich,

123 For some dialects, this labeling makes sense. Within Romani, Sepečides, originally from Thessaloniki but now spoken in Izmir, has *-džar* in the instrumental plural, indicating that it is derived from a true affricate, since the affricate voiced to *dž* (via *-ns-* => *-nts-* => *-nt^{s}-* => *-nd^{z}-*, further with palatal realization).

124 And there are languages, in the Balkans even, that contrast biphonematic [ts] with monophonematic [t^{s}]; in the Greek dialect of Lesbos (Newton 1972: 213–214) [ts] from the loss of an unstressed high vowel (thus [. . . tis . . .] => [. . . ts . . .]) contrasts with a true affricate [t^{s}], e.g., [kat^{s}i] 'sit!' (StGrk κάτσε) vs. [matsi] 's/he plucked' (from earlier [maðise], cf. StGrk μάδησε). Cf. also Pol *trzy* [tši] 'three' vs. *czy* [či] 'whether,' or the classic English example *catch it!* [kæčIt] vs. *cat shit* [kætšIt].

125 See §5.4.6 for discussion, based on Joseph & Tserdanelis 2006 of subtle differences in stop-plus-sibilant combinations in the Balkans based on measurements of the duration of each part of the combination.

who in his 1862 work (p. 6) refers to the loss of *-l-* before *-i-* in both Romanian (as in /celi/ 'the.PL' becoming *cei*) and Albanian (as in /kardinal-i/ becoming *kardinai* '(church) cardinals.' The particular loss that attracted Miklosich's attention may not be significant – a lateral in a palatalizing or fronting environment may be inherently prone to move in the direction of [j] and thus to suffer possible loss before *i*[126] – but looking at a wider range of facts reveals that the laterals present a set of outcomes across the languages of both a divergent and convergent nature.

One point of similarity but also difference has to do with the status of dark *l* and clear *l* in the various languages. The basic facts are summarized well by Friedman 2006a: 660:

> The alternation of clear /l/ before front vowels and velar /ɬ/ elsewhere is characteristic of B[alkan]S[lavic] (including Torlak [and southernmost Montenegrin] but not the rest of B[osnian]C[roatian]S[erbian]), Northern Greek, Balkan Romani, and Vlah, but not Albanian, where the two sounds are in phonemic contrast, nor Romanian and Southern Greek, where only clear /l/ occurs.

There is more to say on the matter, however. Seliščev 1925: 49, citing Weigand, writes that velar /ɬ/ in Frasheriote and Tirana Aromanian is a result of contact with Albanian. (Cf. Kharalmova 2020, who makes the same argument for Aromanian in Selenica in southern Albania, and see also Neiescu 1997: passim.) Stevanović 1935: 43–45 notes that in most of eastern and southern Montenegro, *l* is automatically clear before front vowels and normally dark before back vowels, as in most Macedonian dialects (see also R. Greenberg 2000; Morozova & Rusakov 2018ab; Morozova 2019, 2021 and references cited therein). This alternating *l* is opposed to the palatal *l* of most of BCMS.

Inasmuch as these languages form a contiguous geographic band in the central Balkans, including East South Slavic (Macedonian and Bulgarian), and given that this alternation appears to be innovative in Northern Greek at least (since the southern dialects that lack it tend to be more phonologically archaic than the north), it seems likely that this distribution reflects a local regional convergence. Significant here too is the fact that the Torlak BCMS dialects preserve a velar lateral in positions where most other dialects of BCMS have vocalized to *o*.

Moreover, telling facts from Balkan Turkish can be added to this description that further support the local regional interpretation. Friedman 2003a: 59 observes that Balkan Turkish diverges from standard Turkish in certain ways with respect to the laterals. Standard Turkish has a clear *l* and a dark *l* that – except for loanwords from Arabic (e.g., *lâzım* 'need') or from French (e.g., *rol* 'role,' with accusative *rolü* with a front form of the case suffix presumably due to the fronted lateral) – are in complementary distribution, with the preceding vowel determining the quality of the lateral. In Macedonian, as described in Friedman 1993d: 255, for the most part, the following vowel matters, with laterals realized

126 Note, for instance, as becomes clear below, that the Albanian diasystem has palatal lateral /ʎ/ that gives [j] in some dialects. Moreover, in contemporary Modern Greek, /l/ is often lost in fast speech before *i*, e.g., [maista] 'yes, certainly' from μάλιστα ([malista]).

as clear before front vowels and *j*.[127] Interestingly, there is a tendency for the local West Rumelian Turkish in North Macedonia to realize a clear [l] in accordance with Macedonian rules rather than Turkish rules. In Kosovo, however, where the contact language for Turkish is Albanian, with a different distribution of clear and dark *l* from that found in Macedonian, the local Balkan Turkish pronunciation of the laterals is more in accord with the Turkish distributional rules. We can also note here Judezmo can have a dark /l/ alongside a clear one, presumably under the influence of Turkish (Bunis 1975: 19).

In fact, Turkish was instrumental in the spread of dark *ɬ* and in preventing the loss of clear *l*. The opposition *l*/*lj* (older *ɬ*/*l'*, modern *ɬ*/*l*) was already disappearing in Slavic in the thirteenth century, as evidenced by confusions of the type <лоубо> (*lubo,* etymological *l'ubo*) 'or' and <лючше> (*ljučše*, etymological *lučše*) 'better' in documents from that period, and it was completely lost in the Prilep-Veles region by the time of the Ottoman conquest (Koneski 1981: 56–57). The influx of Turkish loanwords with clear *l* (automatically before and after front vowels, distinctively elsewhere) prevented the spread of the change and reintroduced clear *l* as a phoneme in Prilep-Veles, e.g., *bela* 'white' vs. *belja* 'trouble,' *biljbilj* 'nightingale' vs. *usilba* 'effort,' *bil* 'was.' This did not, however, reverse the effect of hardening of soft *l* where it had already occurred, e.g., *lugje* 'people,' *nedela* 'Sunday, week' (Koneski 1981: 56).

Still, even with similarity at the phonetic level, there are differences in the phonological status of these sounds. In Macedonian (and Romani dialects in contact with it) the clear/dark distinction is phonemic but with a low functional load (Friedman 2002c: 11). In Albanian, by contrast, the distinction is not only phonemic, but robustly so, with numerous minimal pairs, e.g., *lapa* 'skin.DEF' vs. *llapa* 'porridge,' *lule* 'flower' vs. *llulle* 'pipe.GEN.INDEF,' *djali* 'boy.DEF' vs. *djalli* 'devil.DEF,' *mal* 'mountain' vs. *mall* 'longing; property, goods, etc.'

Albanian in fact shows a rather intricate set of developments involving laterals. As Hamp 2002, drawing on Pedersen 1895a, explains, a single Proto-Indo-European lateral, usually reconstructed simply as **l*, in different phonetic environments, was the source for Albanian clear /l/, with direct reflexes as such in the standard language, for the velarized lateral /ɫ/, which shows a wide range of reflexes in the dialects but is /ɫ/ (orthographic <ll>) in the standard language, and for a palatal lateral /ʎ/, reconstructible for Common Albanian and realized as such in some dialects but as [j] in the standard language. The relevant environments are as follows: the clear [l], realized as such or as slightly palatalized or softer in most dialects but as distinctly palatal [ʎ] in some Arbëresh dialects, in Hamp's words (p. 245) "typically derive[s] from initial **l*, medial clusters **-ln-* and **ll-*, and in many dialects from **l* clustered with an obstruent"; the velarized /ɫ/ comes from

127 As Friedman makes clear, the laterals in Macedonian are phonemic, since clear *l* can occur before back vowels, in part due to loans, which have given minimal pairs such as [beła] 'white.F' versus [bela] 'trouble' (from Turkish). Still, the functional load for this distinction is extremely low.

intervocalic **l*[128] in inheritances from PIE, e.g., *siell* 'bring' from **k^{w}ele-*, but also in Latin loans, e.g., *ullî* 'olive' from Latin *oleum*/*oliuum* – the dialect reflexes of this sound are quite diverse and include [ɬ], [l], [w], [v], [x], and [ɣ];[129] and, the Proto-Albanian */ʎ/ derives typically from **li* or **lj*. There are dialects that distinguish all three as laterals, e.g., the Arvanitika of Sofiko and of Kranidi, and in a sense the standard language distinguishes them by having three distinct reflexes, just not all as laterals, but rather as (orthographic) <l>, <ll>, and <j>.

Besides simply showing how a complex set of reflexes can develop by regular sound change, this Albanian situation is interesting for the Balkans and for language contact in three ways. First, some of the diversity of outcomes for the velarized lateral and the palatal lateral in Albanian dialects is matched in dialects of Aromanian and Bulgarian, respectively, though the facts here are of typological significance only, not of relevance for Balkan linguistic history or language contact. For instance, Caragiu-Marioţeanu 1968: 48 notes that Frasheriote dialects of Aromanian have *ɣ* as a development out of /l/,[130] most likely, she says, a velarized variety, as in [ɣok^{u}] from [lok^{u}], and Sawicka 1997: 32 cites dialectal Bulgarian [zeje] 'greens' from [zeʎe]. Second, even though Arvanitika is dying a slow death in Greece, and has been increasingly moribund for decades, resulting from what Hamp 1989b: 201 has called the "modern intense Hellenization" phase of Arvanitika-Greek contact, some aspects of the phonology can be remarkably conservative and can be retained, even under extreme conditions of contact. Such is the case with the three laterals of Sofiko and Kranidi. Third, Hamp 2002: 249 suggests that the Albanian threefold treatment of PIE **l* and the lateral developments in the language in general are connected, via a substratum type of language contact, with the three-fold laterals in the Torlak BCMS dialects (where one finds clear, palatal, and velarized laterals, cf. above); as he puts it:

> When we recall, from toponymic and Romanian evidence, that the South Slavs who became Serbs occupied an area of former Albanoid speech, it becomes quite reasonable that the resultant population was well equipped to conserve the richest system of lateral distinctions and alternations among the later Slavs.

Another phonological phenomenon involving laterals that is contact-induced is the velarization of clear /l/ after a front vowel if followed by pause or consonant, e.g., [bülbüł] for *bülbül* 'nightingale,' for which Ibrahimi 1982: 55 provides Turkish spoken by Torbesh Muslim Macedonian speakers (in western North Macedonia), who carry over their Macedonian speech habits regarding laterals (with dark variants before consonants or pauses) and thus say [bułbuł], as opposed to [bülbül] in the Turkish spoken natively by Turks in North Macedonia and [bülbül] by Turks in Kosovo.

128 See §5.4.4.9.1 immediately below for Hamp's interpretation of the relationship between Albanian intervocalic developments with **l* and the Romanian change of **l* to *r*.

129 Such a development thus adds to the ways in which a [ɣ] can enter Albanian; see also footnote 115, and, on loanword sources of this sound, §5.4.4.3.

130 As noted for Albanian in footnote 115 (and see also footnote 129), this Aromanian development provides another way in which a [ɣ] enters that language; §5.4.4.3 has more on loanword sources.

One further point about laterals is that besides the developments illustrated here – involving differing realizations of laterals – the loss of laterality with a palatalized or palatal-like lateral must also be reckoned with in some instances, especially after labials, in Balkan Slavic, some Albanian dialects, and some items in Greek. Both Macedonian and Bulgarian lose the Common Slavic epenthetic /l'/ that arose in combinations of labial plus jot, but only at morpheme boundaries, e.g., *zemja* 'earth' from earlier *zemlja* (root *zem-*, with *-ja-* stem formative) vs. root-internal *plju-* 'spit' (from earlier **pju-*).[131] This same kind of change occurs in some Albanian dialects in North Macedonia, e.g., Debar/Dibra, Kičevo/Kërçova, Upper Polog (Gostivar region), and some adjacent regions, although it has also been recorded from some West Central and Northwest Geg points (Gjinari 2007: 105–106; Jusufi 2011: 184–185), e.g., *pjak* <*plak* 'old.'[132] A similar phenomenon also occurs in Greek, but only sporadically, as in the comparative marker πιο from earlier πλέον 'more' (unless it is a borrowing from or influenced by the Romance *pio*, as in Italian *più*). While delateralizations next to labials are suggestive, especially in the case of the Albanian/Macedonian similarities, it is difficult to argue for contact-induced change in the absence of detailed documentation.

A last observation worth making about laterals is that they figure in sociolinguistic emblematicity and stigmatization in some parts of the Balkans. For instance, the "dark *l*" of northern Greek dialects, which are those historically (and to some extent still) in the most intensive contact with Macedonian and also Albanian (and, before 1923, Turkish), where the sound is present, is a flashpoint for dialect differences between Greek speakers in Athens and those in Thessaloniki; it is often associated with a stigmatized working-class pronunciation. Similarly, in Kosovo, the local Albanian *l* has a more palatal quality than the Serbian clear *l*, Albanians preserve their native pronunciation of *l* in speaking Serbian. This pronunciation is emblematic of an Albanian "accent" in the pronunciation of Serbian. It thus identifies native speakers of Albanian when they are speaking Serbian and as far as Serbian speakers are concerned, it is a stigmatized way of pronouncing their language. There may well be other such sociolinguistic values attached to laterals elsewhere in the Balkans but these two are particularly salient.[133]

131 Šekli 2008: 103–104 formulates the rule in terms of initial and noninitial sequences of labial plus /j/. However, given the Indo-European root structure inherited by Common Slavic, noninitial C + j would involve a morphemic boundary in any case, whereas initial C + j would not, and prefixed roots with labial + j also retain the epenthetic /l'/.

132 Another delateralization is found in most of Albanian, including the standard language, where, [k']/[g'] (orthographic <q/gj>) occur as reflexes of earlier *klj* / *glj*, which latter are still found in Çam dialects, as well as Arvanitika and some Arbëresh dialects (Gjinari 2007: 189–192), e.g., standard *quhem* 'I am called' / *gjuhë* 'language' vs. dialectal *kljuhem* / *gljuhë*; northern Geg also loses the /l/, but with resultant /kj/ or /k/.

133 We note that in Macedonian, use of velar [ɬ] for clear /l/ is characteristic of the younger generation of urban speakers, especially in Prilep. A complaint of the Macedonian linguistic establishment during the second Yugoslav period was that orthographic <л> was pronounced [l] in all positions (as opposed to the standard, which requires clear [l] before front vowels and velar [ɬ] elsewhere) and orthographic <љ> was pronounced as a palatal [ĺ] rather than the normative clear [l], owing to Serbo-Croatian influence (cf. Korubin 1976: 107–12, 1980: 157–66). It is worth noting that the rules of the standard, as in most respects, are based on the western central dialects.

5.4.4.9 Rhotics (excluding Rhotacism)

Rhotics offer several sets of facts of potential interest from a Balkanological standpoint in that they show recurring effects across a subset of the languages of the area that have a possible contact basis. The matter of rhotacism – the change of *n* to *r* in Tosk Albanian and some dialects of Romanian – is treated elsewhere (§3.2.7 and §5.4.4.10.5) and thus is not discussed here, though clearly relevant to any full consideration of rhotics in the Balkans.

5.4.4.9.1 *l* > *r* (Grk, Alb, Rmn, dialectal Slv) [FM 7, IS 33–34, PA 39]

This particular feature, noted already by Miklosich, provides a link between laterals and rhotics and refers to the fact that Greek, Albanian, Romanian, and possibly, but only to a rather limited extent, Slavic show a development by which earlier [l] becomes [r] in some environments. According to Miklosich 1879: 212, the interchange of /l/ and /r/ is "*nicht selten*" ('not rare') in Slavic, and most known instances are either not from Balkan Slavic or do not involve *l* becoming *r*, at least not in the same way as seen elsewhere in the Balkans.[134]

As far as adding Slavic is concerned, since this change occurs elsewhere in the family as well, it is thus not even necessarily Balkan. Moreover, for the most part, the environments for the change in Romanian, Albanian, and Greek show some differences, so that except for very similar developments in Albanian and Greek, it is hard to treat these otherwise enticing parallels as anything but independent developments, although there may be an indirect substratum link between Albanian and Romanian. Miklosich himself was properly cautious about this feature, classing it among those on which he would place "lesser importance" (see footnote 78). Still, the facts bear repeating here and can benefit from some further discussion, especially since Miklosich gives just a few forms from Greek and only a very limited set is cited in Asenova 2002: 39, where unfortunately there is a lack of clarity that obscures the overall picture,[135] and readers are left to draw their own conclusions from the unclear data given.

In Romanian *r* is the regular outcome of Latin intervocalic *l*, as shown by Lat *basilica* > Rmn *biserică* 'church,' *solem* > *soare* 'sun,' *caelum* > *cer* 'sky,' *exvolāre* > *zbura* 'soar,' and *nebula* > *negură*, among many others. In initial position or

134 Sporadic instances occur in South Slavic but are not compelling. Miklosich cites OCS *kliknõti* 'to shout' and OCS *krikъ* 'a shout,' but in fact **krik-* and **klik-* are both attested throughout Slavic. Similarly, the name *Gligor* < *Grigor* is broadly attested in Slavic, and cf. Grk γλήγορα for γρήγορα 'quickly.' Sobolev 2005b: 202 records *rusarska* for *rusalska* 'pertaining to Rusalija [the festival after Pentecost]' in Ravna (Provadija region, NE Bulgaria). None of these involve regular sound changes, however.

135 In particular, she gives three examples of an *l*/*r* interchange in Greek that actually show a change of original *r* to *l* (via dissimilation, most likely), e.g., αλέτρι 'plough' from earlier *αρότριον (a diminutive formation based on AGrk ἄροτρον). Such examples point to some interaction between *r* and *l,* as with Slavic *Gligor* (see footnote 134), but they differ significantly from the Romanian case and the Greek instances of *l* to *r* below. This same type of dissimilation occurs in the autonym of Frasheriote Aromanians, e.g., in Bela di Sus (Mac Gorna Belica): *Fãrshãlot(s)*.

before consonants, Latin *l* is retained, as in Lat *laudāre* > Rmn *lavda* 'praise' or *ascultāre* > *asculta* 'listen.' Moreover, in a later period, *l* is retained in Slavic loans, even in intervocalic position, e.g., *lopata* > *lopată* 'shovel,' *kobyla* > *cobilă* 'plow line,' *pola* > *poală* 'lap,' etc. There is one possibly telling exception to this development, *măgură* 'hill, hillock,' if somehow from Slavic **mogyla* 'idem' (from earlier **magūlā*, BER IV: s.v.); if of Slavic origin, it may have been borrowed at an earlier stage than the other Slavic words in Romanian or perhaps it has a different story behind it, maybe even a different source altogether.[136]

In Greek, several examples are to be found of *r* for earlier *l*, e.g., αδερφός 'brother,' αδερφή 'sister' from AGrk ἀδελφός/ἀδελφή, ήρθα 'I came' from earlier ἦλθον, αρμυρός 'salty' from earlier ἁλμυρός, and the proper name Στέργιος from earlier Στέλιος. Even loanwords could be affected, as in τσερνίκι 'a type of boat,' if from Slavic *tšŭlnŭ* (Weigand 1928).[137] In each such instance, the *-l-* occurs next to a consonant; in intervocalic position, earlier *l* is unchanged, as in έλα 'come' from earlier ἔλαυνε, or όλος 'all' from ὅλος. The Greek environment for the delateralization is therefore almost the exact opposite of the Romanian environment, thus precluding any connection between the two.

Nonetheless, there may be a contact dimension to the Romanian *l* > *r* change. Hamp 2002: 245 suggests that an "Albanian substratum" may have played a role in this development. As his discussion makes clear, he has in mind the fact that the Romanian change is only in intervocalic position, and in intervocalic position, as noted above in §5.4.4.8, Albanian shows a special development of early *l*, changing it to the "velarized lateral now written in the standard orthography as *ll*" (Hamp *ibid*.). Thus presumably, for Hamp, the Albanoid population that shifted to Latin and thus created Romanian brought to the language shift a characteristic treatment of intervocalic *l* that, in his view, ultimately led to the Romanian *r* in that position. Since velar [ɬ] in Albanian dialects gives a wide variety of outcomes (see §5.4.4.8 above), it is conceivable that one could have led eventually to *r*, especially since there is as much as a millennium or more between the presumed shift of Albanoid speakers to

136 This word presents many difficulties across the Balkans, showing up with reflexes of Common Slavic **magūlā* in South Slavic (e.g., Blg *mógila*) as well as apparently metathesized **gamūlā* (e.g., dial. Blg *gomíla*), and with *a*-vocalism in other languages (e.g., Greek μαγούλα, Albanian *magulë*), which actually provides a better match for the Romanian vocalism, though not necessarily for the lateral. While speculative, we note the Latin word *magalia* 'huts, tents' as a possible clue to the development of Romanian *măgură*. Joseph 2007a suggests that the root of this word, with its shape-based semantic affinity to 'mound,' could have given a Romanian form *magar-* (vel sim.) with 'mound'-like semantics that might have affected the form that **magula* took in Romanian. See also Cioranescu 1958–1966: s.v. for several other possibilities proposed in the literature – none wholly satisfactory – either ignoring the *r*/*l* problem or also positing unattested (but admittedly plausible) Vulgar Latin forms. See also BER IV: s.v for a complete review of the Slavic and Balkan possibilities.

137 Miklosich also cites Greek μπούρκα, from an earlier μπούλκα, but he gives no gloss or dialect source for it; we assume it is a northern Greek form of Bulgarian *bulka* 'bride' (formally a diminutive of *bul-* 'veil,' itself of obscure etymology). Nothing like it occurs in two of the largest Modern Greek dictionaries (Babiniotis 1998, LKN), though Charalambakis 2014: s.v. has μπούρκα for the dress of Muslim women, but quite rightly considers it a relatively recent borrowing from English (ultimately from Arabic). Thus Miklosich's μπούρκα is probably a nineteenth-century dialectal "occasionalism" (see also footnote 32).

Latin and the first attestations of Romanian (sixteenth century). Still, there is no basis for connecting the Romanian developments with the Greek delateralization.

One form that must be mentioned in this regard is the old ethnonym connected with the Albanians, namely the form that shows up in Greek as αρβαν- (in the Albanian dialect name Αρβανίτικα, spoken by Αρβανίτες), and in Albanian itself as *arban-* (Geg, as in the place name *Arbanasi*), *arbër-* (Tosk, as in the designation for the Albanian of southern Italy, Arbëresh and the Arvanitika autonym for the language, *arbërisht*), and *lab-* (the basis for the designation *Labëria* for a part of southwestern Albania, and forms related to that such as *labërisht* '(in) the dialect of Labëria'). All these reflect an original **alban-*, a form whose ultimate origin Hamp 1994a: 66 calls "simply obscure." The various forms can be easily explained from that starting point, however, and show phonology characteristic of various languages in the Balkans, including the delateralization of interest here. The *lab-* form shows the effects of Slavic phonology acting on an original *alb-*, in that metathesis of **alC-* to *laC-* is regular in Slavic (part of the complex process of the elimination of Common Slavic diphthongs of the type vowel + liquid, cf. Shevelov 1965: 391–421); presumably **alb-* was borrowed into Slavic and then back into Albanian after the metathesis. The *r*/*n* difference between the Geg form and the Tosk form (including *Labër-*) is a reflection of Tosk rhotacism (see below §5.4.4.10.5). As for the Greek form, while Andriotis 1983: s.v. Αρβανίτης, following Fourikis 1931, claims that Greek borrowed the -ρ- form of this ethnonym from Albanian itself based on a toponym *Arbena*, it most likely (so Hamp 1994a: 66) rather shows the Greek delateralization before a consonant, i.e., **alban-* > **arban-*, with the borrowing being early enough so that the *-b-* would develop, as the Ancient Greek sound does regularly, into *-v-*.[138] In that case, the Albanian *-r-* in the first syllable (*Arban-*/*Arbër-*) could reflect a borrowing from Greek, rather than Albanian delateralization per se; it is interesting to note, however, that there are occasional examples of *r* for *l* cited for Tosk – Hahn 1854[2]: 14 gives *bilbil* > *birbil* 'nightingale,' *qelbësirë* > *qerbësirë* 'dirt,' *cilli* > *ciri* 'which?'.

Still, any connection of occasional Albanian delateralizations with the Greek delateralization would be difficult to maintain since the Albanian context is broader than the Greek context, as it includes intervocalic *-l-*. Thus, we follow Hamp's account of *arban-*/*arbër-* as showing Greek phonology, and remain agnostic on the status of Hahn's Tosk forms.[139]

Although more data about the Albanian instantiation of this change might change the overall interpretation, it seems that we are left with merely suggestive hints but nothing substantive to go on with regard to a contact-inspired delateralization throughout the Balkans.

138 The modern *-l-* forms of this ethnonym, such as ModGrk Αλβανο-, are borrowings from Lat *albanus*, itself a borrowing from the old Balkan term.

139 Sh. Demiraj 2006: 167–187, in a long treatment of "the denomination of the Albanians in the course of centuries," discusses all the derivatives mentioned here, and some others, but is noncommittal about the form of the original etymon, saying only that it "might have had its origin from a common noun *arb-* or *alb-*" (p. 174).

5.4.4.9.2 *rn/rl* > *rr* (trill) (Alb, Aro, dialectal Grk(?)) [KS 114–115]

Sandfeld 1930: 114 (and no one since) makes brief mention of a convergence between Albanian and Aromanian that he takes, following Capidan 1923: 499, to be an "*influence de l'albanais sur le phonétique aroumain*" ('influence of Albanian on Aromanian phonetics'). In particular, both in Albanian and in the Aromanian of Frasheriotes of central Albania, as well as the language of Aromanians of North Macedonia originally from Albania, one finds *rn* and *rl* both developing into *rr*, which he describes as "*un r long fortement roulé*" ('a long *r* strongly rolled'). As examples he gives on the Aromanian side *ficiorru* 'the son,' from earlier *ficior-lu*, and *iarră* 'winter,' from earlier *iarnă*, and on the Albanian side *barrë* 'burden,' from Indo-European **bhor-no-*, and two loanwords, *gorricë* 'wild pear tree,' from Blg *gornica*, and *prrar* '(a type of) oak tree,' from ModGrk πουρνάρι 'yew tree,' and he accepts the view of Jokl 1926: 89 that the loanword evidence means that this is a relatively recent change in Albanian.

Capidan provides several examples, but even so, with just these accounts to go on and no other reports of this phenomenon to work from, it is a difficult one to evaluate. It does seem to be a plausible feature for transfer from Albanian into Aromanian in a bilingual context, with the apparently recent Albanian assimilatory tendency involving clusters with *-r-* and the presence of a distinction between *r* and *rr* both being features of Albanian that could easily extend into the usage of Aromanian speakers using Albanian in an Albanophone milieu. Nonetheless, a cautionary note is in order.

In Gorna Belica (Aro Bela di Suprã) in southwestern North Macedonia, on the Albanian border, where, since the arrival of the Frasheriote Aromanians from Albania in the late nineteenth and early twentieth centuries, there has been sustained contact involving Frasheriotes and both Macedonian and the local dialect of Aromanian (Mbãliote), the *r*/*rr* distinction has been lost, and there is no evidence of the assimilatory treatment of *r* + *l* (Markovikj 2007); rather, outright loss of *-l-* occurs in forms like /ficior-lu/, giving *ficioru* (a variant that Capidan 1923: 498 actually mentions). Since this is essentially the same sort of population that Sandfeld's account is based on, though some 100 years later, it is reasonable to suppose that there has been change here. In particular, there could well have been a more recent loss of an earlier Aromanian *r*/*rr* contrast in a linguistic environment involving a language with no such distinction, Macedonian, a development which is discussed further in the next section.

A possibly related phenomenon that is also hard to judge but is suggestive nevertheless is the loss of *-n-* in *-rn-* clusters in the Greek dialect of southern Albania, where πουράρι for 'pine tree,' from πουρνάρι, as above, and present tense forms such as κερώ 'treat' for earlier κερν- (StModGrk κερνώ, from AGrk κεράννῡμι 'mix (wine with water); give to drink') occur. Greek does not have an *r*/*rr* distinction or geminate consonants in general, so that *-r-* would be the expected outcome here if there had been an assimilation of *-rn-* to *-rr-*. This dialect occurs in an area where Greek speakers are in contact with Albanian speakers and are fully

fluent in Albanian, and where some speakers may even have shifted from Albanian to Greek;[140] thus either a (substratum) interference with Albanian phonological patterns transferred into Greek if there were a shift to Greek or a reverse interference effect from Albanian onto speakers whose first language is Greek is a distinct possibility here.

Still, however plausible this localized development might be, more information – which admittedly may be lost to the ages – is needed on it before anything more secure can be said.

5.4.4.9.3 Loss of *r*/*rr* distinction (Aro, Alb, Rmi; Rmn, Jud) [IS 32]

The foregoing discussion is based on the view that under appropriate conditions of language contact, speakers can acquire a new sound and thus a new distinction, in this case between *r* and *rr*. And, it happens also that an existing distinction is maintained, but with different sounds; Sawicka 1997: 32 notes concerning the Albanian *r*/*rr* distinction that "in Geg towns and several Southern towns this opposition is preserved when [R] changes to [r] and [r] into a kind of flap [ɹ]." Such shifts show the loss of the long rhotic segment ([R], i.e., *rr*) but with preservation of the contrast. However, it happens also that the sound is lost and with it the distinction too. Besides the apparent instance of such a twofold loss in Bela di Suprã Aromanian discussed above in §5.4.4.9.2, there are a few other cases in the Balkans in which the loss of this contrast can be attributed to language contact and thus appears to be a local Balkanism or a group of several local Balkanisms.

For instance, according to Gjinari 1989: 185, the long trilled *rr* is lacking in Berat, Dibra, and Korça and in the southern towns in general, while according to Gjinari 2007: 90, the neutralization occurs in Tetovo, Dibra (Mac *Debar*), Resen, Korça, Berat, Vlora, and Delvina. This is at odds with some of Sawicka's claims given immediately above. However, while this might be a purely language-internal development – note that Sawicka 1997: 32 says about Albanian "long [R]" (i.e., <rr>) that it is "an isolated phonem [*sic*], consequently it tends to be lost"[141] – it is more likely to be attributable to Turkish influence, since Turkish, with no *r*/*rr* distinction, was the language of the towns in that region for several centuries. Moreover, the other languages in the multilingual urban centers also lacked such a distinction. A contact explanation, moreover, seems warranted for a similar change under similar conditions among younger (terminal) Arvanitika speakers in Greece (Sasse 1991: 57–58; Tsitsipis 1998: 24), inasmuch as Greek does not have this sound either. Greek, Turkish, or Bulgarian could also account for the loss

140 These observations are based in part on BDJ's fieldwork in the area in 2010–2012; see also Spiro 2008 for a description of the Greek of Délvino (Alb Delvina), one of the towns in the region.

141 See also footnote 61, where some general considerations against markedness-based accounts are given. Dickerson 2020 examines the loss of this distinction in Albanian experimentally, from a sociophonetic perspective, viewing the loss in the context of a general ideology of a north versus south opposition within Albanian society.

of the distinction in Eastern Diaspora Albanian, where at times the trill and the flap are in free variation (Liosis 2021, and sources cited therein). Moreover, the absence of the *rr* in the Albanian of parts of Kosovo and in Debar (Elezović 1950: 65) correlates with the absence of the sound in Turkish and Serbian/Macedonian, other key languages historically in the region.

In Romani too, where there is an original distinction of two types of rhotic, the plain /r/ and a second rhotic that can be a long trill or a uvular [R] or a retroflex, first discussed in detail by Gilliat-Smith 1910/1911.[142] As portrayed in Boretzky & Igla 2004: 2.115 and Boretzky et al. 2008: Maps 3 and 4, it appears that distinction was lost in the dialects of North Macedonia, Greece, Kosovo, and Turkey but preserved in those in Albania, Montenegro, Bosnia-Hercegovina, and Serbia (but cf. Gilliat-Smith 1910/1911, according to whom the retroflex was more widespread than sources indicate now).[143] The dialects in Romania and Bulgaria have mixed results. However, the Vlax dialects tend to be more conservative than the Balkan (in the Romological sense of the term) in this respect, and while the patterns of loss do not follow dialectal lines consistently, at least some of the Vlax dialects in North Macedonia do preserve the distinction (VAF field notes). Moreover, the loss also occurs in most dialects outside the Balkans (Boretzky & Igla 2004: 2.7). In all cases where there is loss, the dialects are in contact with languages without the *r*/*rr* distinction, but dialects with preservation are usually not in contact with dominant languages with such a distinction. Still, it is worth noting that in Bulgaria the loss of this distinction in Romani is on-going (Boretzky & Igla 2004: 2.117). Overall, though, the Romani case is not completely clearcut.

In what might be a relevant development for understanding the Romani situation, the Romanian facts concerning rhotics are noteworthy. Within the history of Romanian, there was a distinction between a "strong" rhotic and a less strong one. Boretzky 1991: 7, citing Rosetti 1968: 522, describes the situation as follows:

> Roumanian once had two qualities of *r* (and two phonemes), a normal apical *r* and a strong, probably more cerebral *ɾ*, written in texts of the 16th century as *rr* or by a special Glagolitic letter. This strong *ɾ* is derived from Latin initial *r* (cf. the Ibero-Romance languages!), but it occurs in words of unknown origin and in loanwords as well.

Contact with Slavic, which did not have such a rhotic distinction, could well have played a role in the ultimate loss of this distinction in the historical period in Romanian; Romanian could then in turn have influenced Romani in this regard.

Thus, it can be argued that in each such case, a high degree of familiarity with and use of a majority language without *-rr-* seems to have been a key factor in the failure of a native contrast to be maintained. The mechanism would thus have been reverse interference, with the interference flowing from the socially dominant second language onto the speakers' native language.

142 Gilliat-Smith, unfortunately, is somewhat imprecise as to the finer phonetic detail of the retroflex.

143 In at least some Džambaz dialects in North Macedonia, the retroflex occurs sporadically (Bodnárová 2018b).

The situation in Judezmo is more complex. The inherited Hispanic contrast between the flap <r> and the trill <rr> is lost in most of Bulgaria (but not Thrace), and all of Romania, Bosnia-Hercegovina, Serbia, and Croatia, but not in Greece, Macedonia, and Turkey (Quintana Rodríguez 2006: 84–88, 111, 376). When viewed as a map (Quintana Rodríguez 2006: 376), it is striking that the regions where Albanian was widely spoken are those where the *r*/*rr* distinction is preserved. Quintana Rodríguez 2006: 111 is explicit in connecting this to language contact.

5.4.4.10 Features Restricted to Albanian and Romanian (including Rhotacism)

There are several features that are restricted to Albanian and Balkan Romance, and in some cases just a part of Balkan Romance, e.g., found in Aromanian but nowhere else. These features are of potential significance in light of the evidence of prehistoric connections between these two languages, as discussed at various points in this work,[144] for they could add a further dimension to the Albanian–Romanian substratum scenario. At the same time, however, if they are more recent and found in Albanian and Aromanian, then they are of interest as potential local Balkanisms. These features, accordingly, are listed and evaluated here.

5.4.4.10.1 *kt*/*ks* > *pt*/*ps* Schuchardt 1868: 49; KS 125, 126–127, and regarding the relevance of *kt* > Alb *jt*, Weigand 1927: 178, note 1, cited in KS 126]

It was recognized relatively early on, by Schuchardt 1868 and by Weigand 1927, that there was a potentially interesting parallel between Albanian and Romanian in their treatment of the first element of earlier clusters of *-kt-* and *-ks-*. In particular, both languages show outcomes with a labial first element. In the case of Albanian, the development is *-ft-*, and it shows up in certain words that seem to represent Latin borrowings, while in Romanian the outcome is *-pt-* and it occurs in words that seem to represent part of the language's Latinate inheritance.

Examples from Romanian include *fapt* 'fact' from Lat *factum*, *opt* 'eight' from Lat *octo*, and *lapte* 'milk' from Lat *lactem*, with this outcome thus falling together with that of original *-pt-* clusters, as in *şapte* 'seven' from Lat *septem*. Examples from Albanian include *traftar*, cited by Hamp 1966 as being from Lat *tract-*.[145] There are even two widely cited words involving the same Latin source form: *coxa* 'hip' > Alb *kofshë* 'thigh' / Rmn *coapsă* 'thigh,' and *lucta* 'wrestling match' (thus: 'fight') > Alb *luftë* / Rmn *luptă* 'battle, war.'[146]

144 See, for instance, §1.2.1.4, §3.2.2.7, §4.2.1.1, and §7.9.2.

145 This word is not glossed, but the meaning 'funnel' is given at <https://sq.glosbe.com/sq/en/taftar>.

146 The semantic shifts seen here probably reflect Vulgar Latin meanings for these words; note, for instance, that outcomes of *lucta* in modern Romance languages (e.g., French *lutte*, Spanish *lucha*) extend beyond the semantic sphere of wrestling.

Complicating the assessment of these facts is the further observation that Albanian has some words ultimately of Latin origin in which a different outcome, *-jt-*, with no labiality at all, is found. For instance, Latin *d(i)rectus* 'straight' yielded Albanian *drejt* 'straight, correct, right,' *tractō* 'drag; manipulate' gave *trajt-oj* 'I treat, form' (note *traftar* above, apparently from the same Latin stem), and *fructus* 'fruit' developed into *fryt-* 'fruit' (with *-uj-* realized as front rounded *y*).

Based on these facts, several interpretations have been proposed, all involving language contact but in different ways, and in each case, different in certain ways from the sort of contact that led to some of the more plausible local Balkanisms discussed in this chapter.

On the one hand, the different outcomes in Albanian could simply reflect two chronological layers of borrowing (as claimed for different outcomes of Slavic **x* in Romanian – see §5.4.4.6 and footnote 122). Presumably, the older outcome is the labial, being ostensibly shared with Romanian, while the nonlabial outcome is Albanian-specific and thus under normal methodological assumptions would be viewed as the more recent development.

With this assumption, a plausible scenario for this apparent shared development can be constructed, working from Hamp's claim that an Albanoid substratum shifted to Latin, forming Romanian. In particular, the relevant period of Albanoid would be well after Proto-Indo-European clusters of **ḱt*, **k$^{(w)}$t*, and **pt* lost their initial segment (note *tetë* 'eight' from **oḱtō-ti-*, *natë* 'night' from **nok$^{(w)}$-t-i-*, and *shtatë* 'seven' from **septṃ-ti-*). At that stage, Latin sequences of *-kt-* in the mouths of the shifting population could well have been altered (adapted) to LABIAL + *t*, since there were no native *-kt-* clusters to serve as a model (though also, admittedly, no *-pt-* clusters);[147] as the shifting speakers continued their use of Latin, and Romanian emerged, *-pt-* would have been the regular sound change development out of Latin *-kt-*. In the population of Proto-Albanian speakers that developed into the Albanian speech community, Latin loanwords with *-kt-* would have been altered (adapted) in the same way, to a labial first element. One has only to assume a further change at some point later in Albanian by which the labial fricativized, giving the attested *-ft-* outcome.

At a later stage, then, after this type of adaptation of *-kt-* clusters took place, other Latin words with *-kt-* would have entered the lexicon and, one has to assume, would have been adapted differently since the relevant sound changes were over and done with and the phonotactics could well have been different. This would be the point at which *-jt-* arose.

While plausible, this is a fairly complicated scenario, and it does not offer a clear explanation for why a labial in particular would have been the outcome. Moreover,

147 Presumably, the shared gravity (using the Jakobson et al. 1967 acoustically based feature – see §5.4.4.6) of [k] and [p] figured in this reinterpretation of [kt] as [pt].

in terms of a source for the forms with the ostensible later outcome, it has been noted (by Hamp 1966, citing Barić 1957) that Old Dalmatian offers a ready source for Albanian *-jt-* words, since it has *-jt-* from Latin *-ct*, as in *traita* from *tract-* (note the Albanian above). Thus the words with the *-jt-* outcome need not be from Latin directly at a later stage but could be from a different immediate source (even if ultimately of Latin origin).

But if that is the case for the *-jt-* words in Albanian, it is fair to ask if the labial outcome must also be directly from Latin. That is, it could well be that LABIAL + *t* reflects the regular sound change development of Latin into Romanian and that the words in question were borrowed into Albanian and developed as seen above to *-fC-*. In this regard it is perhaps instructive to note that in *kofshë*/*coapsă* both Romanian and Albanian show the same semantic shift from Latin 'hip' to 'thigh,' a fact that is consistent with a borrowing relationship.

Thus, the Albanian dual outcome of clusters ultimately from Latin *kt*/*ks* could reflect not different chronological layers of borrowing but different sources, East Balkan Romance (essentially Romanian) and West Balkan Romance (essentially Dalmatian), and is thus what Hamp 1966: 105 calls an "inner Balkan" historical and geographical layer in the history of Albanian and a "coastal Adriatic" layer. Moreover, the direction of the borrowing is not an issue, even though there are not many obvious loanwords from Romanian into Albanian while there are several that went in the opposite direction from Albanian into Romanian, since, as Hamp puts it, "the loan situation may easily be more complex" than Romanian always being the borrower and Albanian always the donor.

Another consideration to keep in mind concerning the labial outcome is that the shift from a velar to a labial in clusters like these is not too surprising a change. On the one hand, the acoustic feature [+grave] links labials and velars (as noted in §5.4.4.6, and cf. footnote 147), so that the substitution of one for the other passes the test of phonetic naturalness.[148] On the other hand, a stop consonant is less likely to be fully released in a cluster and an unreleased *-k-* could well be perceived as, and thus ultimately changed to, an acoustically similar but articulatorily different stop, especially if, as here, loanwords are at issue. In that case, one might be tempted to say that the Albanian and the Romanian developments did not have anything to do with one another and the early Albanian labial forms could be just the result of an adaptation of loans at an early period.[149]

148 There are some labial/velar confusions in Balkan Romance, e.g., Aro *kizda* < Slv *pizda* 'pudendum mulieris,' dialectal Rmn / Aro *gine* 'good, well' < *bine* (and elsewhere in Romance too – Subak 1906: 160 mentions nonstandard Sp *gueno* for more usual *bueno* 'good'). Since these are sporadic and since labials and velars are acoustically linked, these are most likely unrelated to the Latin *-kt-* outcomes discussed here.

149 Somewhat more speculatively, if the Albanian labial outcome is indeed independent of the Romanian one, the path of development could have been *kt* > *xt* (a change found in later Greek, e.g., οχτώ 'eight' from earlier ὀκτώ), and then later to *ft* via the same sort of change seen dialectally in Albanian with *-h-* (see §5.4.4.6, but with some reflexes even in standard Albanian). If so, and this account certainly has its own problems, then the Albanian developments would be even more removed from the Romanian ones than generally admitted.

In the end, it is hard to tell what to make of this parallel. The evidence would seem to be more in favor of viewing it as not particularly indicative of a special connection between Albanian and Romanian in an early period, however that relationship is to be envisioned. To be sure, language contact is involved in these developments, temporally quite far removed from what we have characterized as the critical period of Balkan sprachbund formation, though certainly not uninteresting. But the particular type of language contact involved may well be simply a rather trivial kind of loanword phonology, and nothing more.

5.4.4.10.2 *b* > Ø /V__V [KS 125]

Sandfeld 1930: 125 states that an "*évolution commune*" ('common evolution') between Albanian and Romanian appears to be called for based on cases like Rmn *cal*, Alb *kal* 'horse' both from Latin *caballus*, and Rmn *cot* (from *cuàt*), Alb *kut* 'unit of measure, ell, elbow' both from Latin *cubitus* 'elbow.' The shared loss of intervocalic *-b-* is striking, especially when lined up with other convergences suggestive of an ancient link between these languages. This one, however, is not as significant as might seem at first glance, and its probative value is suspect.

The main problem here is that there are not many examples on the Romanian side and overall they suggest a somewhat different context for the sound change from what is evident in Albanian. For one thing, *cot*/*kut* could show some influence from Greek κουτί 'box' (i.e., as a container with elbow-like corners), a diminutive of Ancient Greek κύτος 'box.' Still, there are other examples, including *sulă* 'awl' from Lat *subula* and *staul* 'stable' from Lat *stabulum*. More importantly, though, if *lucra* 'to work,' from Lat *lucubrare* 'work at night,' belongs here, then it would seem that in Romanian (or, better, Balkan Romance), the change was loss of word-medial *-b-*, not intervocalic *-b-* per se. And even so, a few problematic forms also exist as potential counterexamples, such as *zgaibă*, discussed below in §5.4.4.10.4, assuming it is from Lat *scabies*, and possibly *negură* 'cloud' from Lat *nebula*.[150]

As for Albanian, the input to the change has a wider scope than in Romanian, taking in all voiced stops at a certain period. The effects of this change are especially evident in loans from Latin, e.g., *mjek* 'doctor' from Lat *medicus* 'doctor' or *gjyk* 'law court' from Lat *iudicium* 'legal proceedings; court,' *mjeshtër* 'master [craftsman]' from Lat *magister* 'master,' but voiced stops in inherited words from Proto-Indo-European were also affected, as in e.g., *det* 'sea' from **dubeto-*, lit., 'the deep (thing)' or *sot* 'today' from **ḱjā-dīti-* (Beekes 1995: 263; Joseph 2013a). While these examples all involve intervocalic voiced stops, Albanian offers some examples that suggest a broader, word-medial, conditioning too, instead of merely intervocalic, e.g., *erë* 'smell, wind,' if connected to Latin *odor* and thus from**ōd-r-o-*, or

150 There are other possibilities for accounting for the *-g-* of *negură* (see Cioranescu 1958–1966: s.v.), including hybridization (contamination) involving Slavic **mьgla* 'fog' (see Joseph 2007a). Note that *negură* is a problem here only if the apparently dissimilatory change to *g* next to *u* was later than *b*-loss.

ujë ‘water,’ certainly from the **wed-* root of AGrk ὕδωρ and Eng *water* and thus presumably from **ud-r-j-*. The Albanian change has been argued by Orel 2000: 78–80 to be subject to a “rhythmic rule” (cf. also Çabej 1961: 141–142 on this) and to depend on syllable count, affecting only medial voiced stops in words of three or more syllables, which could further differentiate it from the Romanian change. Also participating in some way in this change, to judge from the evidence of some words at least, were the Albanian outcomes of the PIE voiced aspirated stops, as in *ve* ‘widow’ from **widhew-*.[151]

Overall, therefore, given these differences and uncertainties, it seems best to separate the Romanian and Albanian developments here. Moreover, if the change in Romanian is essentially restricted to *cal*, that word might best be taken as a borrowing from Albanian rather than a native Romanian development.[152]

5.4.4.10.3 *w* > *b* [KS 125]

Sandfeld notes this apparent convergence – what he terms “*une évolution commune*” (‘a common evolution’) – between Romanian and Albanian, citing *corb*/*korp*, respectively, ‘raven’ from Latin *corvus*. Both languages seem, therefore, to have developed a (labial) stop out of a Latin (labial) semi-vowel. This development appears to be restricted to just this one word (at least it is all that he, or anyone else, cites). Most likely, however, this is not an indication of any shared characteristic of the two languages, and in fact it is unlikely that either one turned a glide into a stop in borrowing or inheriting this word, as the case may be. Rather, there is evidence elsewhere in Romance for a stop in this form, e.g., French *corbeau* ‘raven’ (the source of Eng *corbie*), suggesting that the *-b-* had already developed in Late Latin.

151 We assume here that *vjedh* ‘steal, rob,’ if ultimately from PIE **wedh-* (the semantics are not overly compelling, cf. Skt *vadh-* ‘strike, destroy’), has a somewhat different history (perhaps with a suffix protecting the **dh-* or a syllable count that exempted the medial stop), yielding the Albanian fricative. Similarly, in *hedh* ‘throw’ from **sḱeud-* (cf. Eng *shoot*), one can assume that the **d* must have been word-final at some point, e.g., in a root-present or root-aorist formation, so that the **-d-* could remain (so Hamp [p.c.]). Moreover, we assume further that the Albanian fricative from PIE **ǵ* and **ǵh* (as in *lidh* ‘bind’ from **liǵ-* (cf. Lat *ligare* ‘to tie’) or *udhë* ‘way’ from **uǵh-o-* (with the root of Skt *vah-*, Lat *veh-* ‘convey’) developed before the loss of intervocalic voiced stops (unless syllable count played a role here too).

152 Sandfeld 1930: 125, note 1 points to a form in Kretschmer 1905, a description of the Modern Greek dialect of Lesbos, namely *kal’tsevgu* ‘I ride a horse,’ that corresponds to a form καβαλικεύω found elsewhere in Greek. This Lesbos form thus also shows loss of an intervocalic labial obstruent in the word for ‘horse’ deriving from Lat *caballus* (Grk β ([v]) here is regular for Lat *b*). Since he mentions it only in a footnote, Sandfeld presumably thought it merely an interesting curiosity without any Balkanological significance. We agree with that assessment, and note further that καλικευ-, the *v*-less contracted form of καβαλικευ-, is found in Medieval Greek twice, in *Pikatorios* (l. 365) and in *Istoria Ebraiopoulas tis Markadas* (l. 307), though this distribution is hardly revealing: *Pikatorios* is a Cretan work but the provenance of the *Markada* is less clear, making Balkanological inferences difficult. Moreover, as Nick Nicholas has insightfully pointed out to us, there is a verb, καλιγών-, with a related meaning, ‘shoe a horse,’ that has a variant καλικων-; this verb derives from Latin *caliga* ‘soldier’s shoe’ (so that the *-g-* form is primary) and in principle could be the source of καλικευ-; in fact, the variant καλικων-, with an unexpected medial -κ-, might show the influence of καβαλικευ-, suggesting mutual interaction between the verbs ‘shoe a horse’ and ‘ride a horse.’

If so, then Albanian presumably borrowed the word with the *-b-* already present, and in the same way the *-b-* form entered Romanian. There is thus no convergence here at all and Sandfeld's citing it as such is misleading.

5.4.4.10.4 *sk* > *zg* [KS 125]

Sandfeld mentions this development in Romanian and Albanian as yet another "*évolution commune*" ('common evolution') to reckon with in the Balkans, but here too the facts suggest something else. Sandfeld cites just one word, Romanian *zgaibă* 'bleeding ulcer, carbuncle,' Albanian *zgebë* 'scabies, itch,' from Lat *scabies* 'eczema, rash, scab'; it also occurs in Aromanian as *zgaibã* 'wound, abrasion, roughness of skin.' This may be the only word to show this development, though Sandfeld is not explicit on this point.

Meyer 1891: 484 is the first to mention this word and this development, and he offers a wider range of relevant forms that Sandfeld would have done well to pay attention to. In particular, Sandfeld's Albanian *zgebë* should probably be *zgjebe* (the final vowel is confirmed also in Newmark 1998: s.v.) and there are several variants attested: *zjebe*, *dzjebe*, and *sqebe*. The variation on the Albanian side is unsettling, to be sure, and may suggest several chronological layers of borrowing or possibly taboo-like deformation due to the unpleasant semantics. In addition, one has to be suspicious here since other instances of Latin word-initial *sk-* give a different result in Albanian, namely *shk* ([ʃk]) as in *shkallë* 'ladder' (from Lat *scala*), *shkëmp* 'seat, throne' (from Lat *scamnum*), and possibly *shkruaj* 'write' (if a borrowing from Lat *scribo*). Moreover, early **b*, from Proto-Indo-European, disappears medially (see above §5.4.4.10.2), and so too in Latin borrowings, as, presumably, in *shkruaj*; thus the retention of the medial *-b-* in this form is unexpected if a(n early enough) loanword from Latin. Complicating the picture is the occurrence of another word in Aromanian with similar reference to bodily disfigurement, *zgrob* 'hunchback,' taken by Papahagi 1974: s.v. to be from (dialectal) Greek [žgróbos], supposedly of unknown origin but cf. BER I:*grăb*, which cites dialectal Grk *γκριμπός* 'hunchback' as well as Balto-Slavic cognates; cf. also Mac *grb* 'back,' *grbav* 'hunchback' with vocalic /r/.

Overall, while the matching oddity of *zg* across the two languages is indeed striking, there is too much about the development of this word that is left unexplained – the voicing to *zg*, the different treatment from other Latin *sk* words, the variation in Albanian – for it to be valuable for any hypothesis concerning a shared Albano-Romanian development. Moreover, especially if *zgrob* is taken into consideration, one can speculate that tabu deformation is responsible for (some of) the peculiarities of the attested forms in one or both languages. It may well be a borrowing from Balkan Romance into Albanian, but it is a form which demands that clarificatory light be cast on it rather than one which offers any illumination on its own.

5.4.4.10.5 *n* > *r* /V__V [FM, KS 126–127, EB, IS 34, PA, SD][153]

This feature, first noted by Miklosich,[154] and picked up by others as well,[155] is discussed at some length in §3.2.2.7, where the purpose is a methodological one having to do with the importance of chronology and taking "Stammbaum" (family tree) relations among languages seriously. The discussion here builds on that in the context of further assessing the viability of this feature as a phonological Balkanism, of necessarily restricted scope, being found in Albanian and Romanian only. More properly, though, as noted in §3.2.2.7, it is not Balkan Romance per se that shows any sort of rhotacism but rather a subset of North Danubian Balkan Romance, in particular Istro-Romanian, and regional dialects of Romanian. And, in Albanian, it is just Tosk that shows rhotacism, as this development is one of the major isoglosses dividing Tosk and Geg within Albanian.

It is clear from the earlier discussion that the distribution of intervocalic rhotacism within Balkan Romance and its distribution within Albanian show that this feature cannot be part of whatever Albanian and Romanian or Balkan Romance inherited from their respective proto-languages, even if those proto-speech-communities were in close contact with one another. It would be hard to construct a plausible account in which the feature entered these languages at that time, possibly through a portion of a substrate population shifting to each one, but only surfacing dialectally within each tradition centuries later.

The only way that the distribution of this feature could be the result of language contact, and thus be Balkanologically significant in the sense developed in this chapter, would be if, as noted in §3.2.2.7, the speech community that became (Southern Albanian) Tosk and the relevant portion of the Balkan Romance speech community (tentatively labelled "Northwest" Romanian earlier) were geographically adjacent and were the only sectors of the Romano-Albanoid area affected by this change. In such a view, it would not have been a substrate, necessarily, that was responsible but rather the change could be seen as an "organic" one in one speech community that was adopted by the other, presumably reflecting bilingualism in that particular region. However, there is no evidence at all for such a geography for these proto-speech-communities, and it is moreover at odds with other dialect linkages within Balkan Romance and within Albanian.

It seems, therefore, that the best assessment is to treat the rhotacism in Balkan Romance and the rhotacism in Albanian as completely unrelated and historically independent (though admittedly parallel) phenomena, despite the initial appeal of linking them. This is the view taken by Sh. Demiraj 2004: 94, drawing on Çabej

153 HS mentions rhotacism (p. 79) but only in the context of features of Albanian dialectology, not in terms of a possible connection with any Romanian development.

154 Miklosich (p. 7) actually talks not of rhotacism per se but rather of a "*Wechsel zwischen n und r*" ('(inter)change between *n* and *r*'); still, the potentially Balkanologically interesting aspect of the interchange is the rhotacizing *n*-to-*r* development (and that is what other scholars have taken up and commented on).

155 Thomason 2001: 108, for instance, lists rhotacism among the "less widespread phonological Balkanisms."

1979: 56, by Rosetti 1924, 1985: 268–269, by Sawicka 1997: 34, who says that rhotacism is "somehow connected with the generally unstable character of sonants," and, though with some equivocation, by Stankiewicz 2002: 369, and it seems to be an inescapable conclusion, given the facts. We note further that although this change is generally referred to as rhotacism, focusing on the segment that results from the change, it can instead be characterized as a "denasalization," with *r* being seen as an *n* that has lost its nasality. In that view, it becomes attractive to see the Tosk Albanian change of *n* > *r* as part of the same process that led to the denasalization of Common Albanian nasal vowels in that dialect (as seen in Geg *âsht* '(s)he is' versus Tosk *është*), as has been emphasized by Eric Hamp.[156] Implicit in such a linkage within Tosk is a separation of the Albanian *n* > *r* development from the Balkan Romance one, as the former has a different, broader scope.

Finally, from a phonetic standpoint, once one accepts the premise that the *n* > *r* change involves denasalization, rather than movement towards a rhotic "target" per se, then two other reasons for separating the two *n* > *r* developments emerge. First, at least some instances of *n* > *r* in the Romanian sphere may be the result of distant conditioning effects; for instance, *fereastra* 'window,' from earlier *fenestra*, may be a distant assimilation, and Istro-Romanian *mâra* 'hand' (Rmn *mână,* Lat *manus*) and *mâre* 'tomorrow' (Rmn *mâine,* Lat *mane*), etc. (Sârbu & Frățilă 1998: 227) may reveal a dissimilatory pressure from the nasal onset in the preceding syllable. Actually a good many of the forms, ten out of fourteen, to be precise, that Miklosich 1862: 7 cites in his brief mention of rhotic developments fall into this category of possibly being dissimilations rather than simply intervocalic rhotacism, so there is less evidence to work with than his more extensive list might suggest at first.[157] No such effects are to be found on the Albanian side, where intervocalic position alone is relevant. Second, if denasalization is involved, then the shared outcome of *r* is not so surprising, given the articulatory similarities between *n* and at least some kinds of rhotic sounds. That is, on naturalness grounds, the parallel between Romanian and Albanian becomes less striking.

We therefore do not judge rhotacism to have any Balkanological significance other than being a major isogloss within Albanian and a sporadic feature within some of Balkan Romance, but we acknowledge, following Trummer 1976, 1981 and Hamp 1981–1982, that the wider denasalization process that it might reflect could well be more significant as a vowel development, discussed below in §5.4.5.1.

156 On linking Tosk "rhotacism" with the Tosk loss of vowel nasalization, see Hamp 1981–1982, working with and extending the insight of Trummer 1976, 1981 on shared processes of denasalization among Albanian, Slavic, and Romance. Cf. also Friedman 2018a.

157 Moreover, three of them actually involve original *r* becoming *n* and thus reflect "anti-rhotacism," as in *sănin* 'calm' from Latin *serenus* 'clear, calm,' where assimilation may be the operative process. Rosetti 1924: 20 explicitly labels some of these examples, and a few others like them, as "*phénomènes d'assimilation consonantiques*" ('consonantal assimilatory phenomena'). Only *suspin* 'sigh' in Miklosich's list, from Latin *suspiru-* 'sighing, breathing,' seems to show "unprovoked" anti-rhotacism.

5.4.4.11 Dispalatalization or Dejotation in Romani

Boretzky 1996: 4–5 suggests that the tendency to dispalatalize palatal liquids and the palatal nasal ([ɲ]), from consonant+/j/, in some Arli Romani dialects of North Macedonia reflects the same tendency in much of Macedonian and could be due to language contact, e.g., *rakla* (< *raklja*) 'girls,' *[h]inum* (< *[h]injum*] 'be.1SG.AOR'), *bojra* (< *borja* 'brides'). While this tendency is quite old in Macedonian (the earliest examples of the hardening of inherited /Ĺ, ŕ, ń/ to /l, r, n/ are late medieval), in at least some regions it appears to have been still on-going at the time of the arrival of the Ottomans and – again in at least some regions – might have continued into the modern period (Koneski & Vidoeski 1983: 50–51, 120). Thus, for example, in southwestern Macedonia, new clusters of the type /nj/ resulting from the loss of weak jers (from /nьj/) were metathesized to /jn/, e.g., *oranьje* > *oranje* > *orajne* 'ploughing.'[158]

5.4.5 Other Possible Phonological Balkanisms

In this section, by way of rounding out the treatment of putative segmental Balkan convergences, we discuss a random assortment of features that have either not been addressed at all in the literature or have not been sufficiently appreciated to warrant mention in the handbooks. Not all are of Balkanological significance but each one does deserve some attention.

5.4.5.1 Vowel Denasalization and Schwa (Mac, Blg, dialectal Alb)

In §5.4.1.6 above, the ubiquity of stressed schwa in the phonologies of Balkan languages, a feature long noted and remarked upon, is discussed and its validity as a Balkanism across all or even most of the region is called into question. We are not discounting the possibility, however, that some schwa developments might be significant on a more localized level, involving a more narrowly defined grouping of languages that are in direct contact with one another.

There is one such aspect of the development of schwa that is noteworthy, and although it has not found its way into any handbooks, there is some relevant literature. In particular, while in all of South Slavic, the Common Slavic back nasal vowel **ǫ* is denasalized (with a few Macedonian dialects segmentalizing the nasality, as discussed in §5.4.1.1, iv), there is a division within South Slavic, noted by Trummer 1983 (see also Trummer 1976, 1981 and Sawicka 2000b, 2014), between languages/dialects with a rounded outcome of the denasalization, as in BCMS and Slovene, i.e., West South Slavic, and those with an unrounded outcome, as in most of the East South Slavic area, what he refers to as the "*u* : non-*u*" isogloss (although the Slovene outcome is generally some form of /o/). Trummer notes further that this isogloss generally matches the Geg–Tosk division in Albanian, one

158 Bulgarian also had such hardening, e.g., *orane* 'idem.'

significant feature of which is denasalization (in Tosk), so that he speaks (p. 251) of the "*non-u-Isoglosse und die gegisch-toskische Hauptisoglosse*" ('non-*u*-isogloss and the Geg-Tosk main isogloss') as being "*eine Isoglosse*" ('*one* isogloss'). Importantly here, the Tosk outcome of denasalization is generally a schwa, as in *është* 'is' (versus Geg *âsht*), so that Trummer's observation becomes relevant for the matter of stressed schwa too.[159] Hamp 1981–1982: 782 (see also 1983: 253) makes the further important observation that this isogloss more or less coincides with the Jireček line (Jireček 1911), and thus may reflect borders of an old culture area; as he puts it:

> I see the unity between the Tosk and Slavic (before Trummer, the inclusion of Romanian had not occurred to me) denasalization as following in remarkable fashion the ancient Jireček line . . . . I see the Tosk and Slavic unrounded denasalization not as 'entwicklungstypologisch(e)', but as a (like Trummer) single (but unlike Trummer) historically engendered idiosyncratic contact phenomenon no doubt resting on the extinct continuum of Roman times which defined by its margin the epigraphic Jireček line.

Further, the usual Bulgarian outcome of Common Slavic **ǫ* is the mid-central schwa-like vowel, while in Macedonian, the most common outcome of that vowel is low central [a].[160] A few Macedonian dialects however have schwa as the outcome, especially within what Friedman 1993d: 301 calls the "Peripheral" (western) dialects of Ohrid and Kostur (but see below),[161] which are within the contact sphere of Tosk. Thus both these Macedonian dialects and Tosk Albanian underwent a denasalizing schwa-creating process, very suggestive of the sort of localized regionalism seen by now with so many phonological processes in the Balkans. Still, it may be that the chronology makes it difficult to maintain this interpretation, since as noted in §5.4.1.1, iv, nasalization may have persisted somewhat late in Macedonian even though the effects of denasalization show up as early as the thirteenth century, but the Geg–Tosk division is a very old one within Albanian and thus may well predate the relevant Slavic developments here. Nonetheless, Hamp 1981–1982: 784 optimistically writes: "The Tosk change of $n > r$ (or its replication in somewhat later loans) cannot be dated very late, i.e., not after the dispersion of the southern Tosk dialects at the turn of the first millennium, at the latest; the same dating should apply to the Slavic narrowed unrounded denasalization which Trummer has discussed." Moreover, as Lindstedt 1988,

159 The actual phonetic realization of Tosk <ë> from a denasalized vowel varies quite a bit dialectally, with not all dialects having [ə] proper as the outcome and some going as low as [æ].

160 In some dialects (see Friedman 1993d: 301) the result is a back vowel [o] or [u], which runs counter to Trummer's isogloss. Hamp 1981–1982: 782 is not worried by this, noting "I am not at all sure there is a true discontinuity between the western extension of the *u*-limit west of Skopje and the Struško Tosk terminus of the Geg–Tosk isogloss, especially if we take into account the incidence of *u*-forms (*put*, etc.) in the Tetovo 'corner,' the Debar *pot* phenomenon, and the sparse settlement in earlier time of the Crn Drim/Drin i zi." We can add here that Debar town Macedonian has schwa, while the rounded vowel of the countryside corresponds to the rounding in the local Albanian reflex of *â*.

161 So also the more westerly dialect area of Ser-Nevrokrop and as far west as Voden-Kukuš, where contact with or influence from the direction of Bulgarian may be involved.

2016 has argued, denasalization may have begun at the southern margins of South Slavic by the ninth century, or even earlier (cf. also Friedman 2018a).

This last scenario is arguably represented in the extreme southwest of Macedonia, basically the Kostur (Grk Kastoria) region, which forms the southwesternmost extension of Modern Macedonian, bordering on and overlapping with Greek to the south, Albanian to the west, and Aromanian co-territorially. As has long been noted (e.g., by Illič-Svityč 1962, and references therein), the Macedonian dialects in this region preserve traces of nasality not found elsewhere in the Slavic world except Polish.[162] Thus, for example, Common Slavic **zõbŭ/zõbi* 'tooth/teeth' is realized along the western and southern edges of Macedonian in the Kostur region as *zəmp/zəmbi, zåmp/zåmbi*, or *zɔmp/zɔmbi*, with the rounded reflexes limited to the southernmost edge. As one moves east or north from the extreme periphery, the nasal is lost before the voiceless stop of the singular but still retained before the voiced stop of the plural. Further still, both remnants of nasality and rounded reflexes are lost, and schwa ends up competing with /a/, the west central (and standard) Macedonian reflex of the back nasal, which has clearly been expanding southwestward even during the course of the twentieth century.[163] Until recently, the homorganic nasals of southwest Macedonian were interpreted as archaic retentions (e.g., in Illič-Svityč 1962). More recently, however, it has been argued that this is really an archaic innovation, i.e., a Balkanism (Friedman 2018a; Lindstedt 2016; Sawicka 2014). While Sawicka 2014 argues for Greek as the source for the innovation, Friedman 2018a argues for a combination of contact factors that include Albanian and Aromanian, as all three languages developed prenasalized stops in various positions (see §5.4.4.1) and lacked or eliminated (in the case of Tosk Albanian) nasal vowels (see §5.4.1.1, iv). Moreover, in Kostur Macedonian, the vowel that develops next to the nasal sonorant occurs nowhere else in the dialect except next to the reflexes of Common Slavic vocalic /r/ and /l/, and in every case the vowel that develops next to the nasal is identical to that which develops next to the former vocalic liquid. This appears to indicate a contact scenario according to which the speakers of languages with neither nasal vowels nor vocalic liquids encountered speakers of a language with both, and in the course of multilingual accommodation (a sort of convergent "drift"), the language of the speakers with the nasals and vocalic liquids were influenced by the pronunciation of the speakers of the languages that had neither. Given the timing of the Slavic invasion of or migration to the Balkans, speakers of southern late Common Slavic and southern late Common Albanian could well have been in contact during

162 Lindstedt 2016, citing Małecki 1934–1936, gives evidence that the border region of the decomposition of nasals extended eastward beyond Thessaloniki (Małecki cites examples from Suho (Grk Sókhos) and Visoka (Grk Óssa), northeast of Thessaloniki), and, moreover, he argues on the basis of the structure of the Glagolitic alphabet that this change was already in place at the time of Cyril and Methodius (ninth century). See also Lindstedt 1988.

163 Nonnasal rounded reflexes appear again north of Little Prespa Lake, and rounded reflexes occur in the Debar region, where contact with Albanian dialects in which *ã* gives nasal or nonnasal *å* or *ɔ*, provide a parallel suggestive of contact. South of Trummer's 1983 rounded reflex isogloss, the only other dialects with rounded reflexes are some Rhodopian Bulgarian dialects.

a crucial period of change and accommodation. Added to this were the pronunciation habits of the local Greek and Romance speakers. A striking feature of the peripheral Kostur reflexes is that the rounded vowel for decomposition is only where contact with Greek – which lacked stressed schwa – was strongest, while the dialects that developed schwa plus nasal or liquid were in more intense contact with Albanian and/or Aromanian, which had that same set of developments (Friedman 2018a).

Thus, a link between denasalization and the emergence of schwa in this one corner of the Balkans cannot be rejected, especially when one adds in the evidence of other denasalizing localized parallels, such as the denasalization with the same outcome in the Geg of the region of Debar located in what was a Macedonian-dominant area, as well as denasalization in the Northeast Geg dialect of southern Montenegro discussed in §5.4.1.4 above.

5.4.5.2 Word-Final Devoicing (Blg, Mac, Mtn, Trlk, Trk, Alb, NGrk, Megl, Aro, Rmi)

Most branches of the Balkan languages show word-final devoicing of stops and affricates. Interestingly, this feature is not mentioned in any treatment of Balkan phonology, other than the brief but extremely useful discussion in Friedman 2006a, yet it arguably has considerable Balkanological import.

The distribution of word-final devoicing is as follows. Within South Slavic, it is found in Slovene as well as Bulgarian and Macedonian (including the Goran dialect of southwesternmost Kosovo and adjacent parts of Albanian), but in BCMS, usually only in border zones such as southernmost Montenegrin and eastern Torlak dialects – as well as in Đakovica (Gjakova), Prizren and along the southern border of Kosovo in Sretečka Župa and Sirinička Župa – where it is sporadically realized with devoiced stops alternating with lax, partially voiced ones (Broch 1903: 45–46; A. Belić 1905: 240–241; Stevanović 1935: 56; Remetić 1996: 442–444 (who connects the phenomenon with language contact); Pavlović 1939: 130–132; Mladenović 1990: 25, 41, 2001: 232–233; Toma 1998: 118; R. Greenberg 2000; Ivić 1991: 104–106; Sobolev 1994: 162–165). It is absent from almost all the rest of BCMS.[164] In South Montenegrin, the occurrence of devoicing is likely influenced by Albanian, whereas in Torlak BCMS, it is a feature shared with Bulgarian and Macedonian.

164 It also occurs in some northern Istrian dialects on the Slovenian border, in Mostar Bosnian, a small part of southwestern Bačka (Serbia) and Galipoli (Pehčevo, North Macedonia), and Krašovan (Romania). The Istrian dialects border on Slovene, and older contact with Turkish could explain Mostar and Gallipoli. Bačka might be due to contact with Slovak or German, but since Krašovan is spoken in Romania, unless there were some German source, it might be an independent innovation (Ivić 1991: 104–106 and references therein).

Final devoicing occurs in Turkish generally but is more consistent in West Rumelian Turkish than in other varieties of Turkish, and is even extended to fricatives, a class of sound not devoiced finally elsewhere in Turkish.

Some Romani dialects also have final devoicing. For instance, some Arli dialects in North Macedonia have it; e.g., /dad/ = [dat] 'father' (cf. /dadeske/ 'father.DAT') is fairly common. Final devoicing of /v/ can sometimes occur, but not with consistency, e.g., *te daf tut* 'that I give you' but *te pučhav tut* 'that I ask you,' both from the same speaker in a short stretch of text (Cech et al. 2009: 206). In the SVlax dialect of Agía Varvára (Greece) there are a few lexical traces (e.g., *yak* 'fire,' with plural *yaga*) of what may have been a more widely realized final devoicing as well (Messing 1988: 14), inasmuch as these speakers came to Greece from Turkey and were bilingual in Turkish in the generations up to 1923.

According to Byron 1976: 95–98, final devoicing is found in Albanian, but only in Northern Geg and Northern Tosk. According to Gjinari 2007: 112, however, the picture is more complex. Some Lab and Çam (southern Tosk) dialects also have consistent final devoicing, as does East Central Geg, while Northeast Geg either lacks final devoicing or only has it facultatively (but cf. Ajeti 1998: 98–99, who writes that it does occur in Kosovo). Northwest Geg favors final devoicing. In southern Geg the distribution of consistent final devoicing, facultative final devoicing, and no final devoicing favors devoicing toward the east (where Macedonian is more prevalent) and preservation of voicing toward the west (where there are fewer Slavic speakers, and those who have migrated there are from BCMS regions where the phenomenon does not occur). Thus, while Byron's generalization is useful, the detailed data presented by Gjinari give a better sense of the facts on the ground.

Within Balkan Romance, final devoicing is mostly absent, except in five of seven surviving Meglenoromanian villages in North Macedonia and northern Greece. All of the relevant area was once Slavic-speaking, specifically East South Slavic – a branch with final devoicing – and there was continued contact with Macedonian after the arrival of Romance speakers in this particular region. Final devoicing also occurs in some Aromanian dialects, mostly Grămostean in Pirin Macedonia (Blagoevgrad [Gorna Džumaja], Razlog [Mehomija], Budova, Popovi Livadi [Papaz Čair] (Aro Papaceair), and Lopovo in Bulgaria) as well as the Gramostean dialect of Ubovo in eastern North Macedonia and in Beala di Ghios (Mac Dolna Belica) near Struga in southwestern North Macedonia, as well as in Gopesh, and, sporadically, also in Molovishte (Saramandu & Nevaci 2014a: Map 93).

Finally, for the most part, final devoicing is not found in Greek, but it is important to note that as an inherited trait from Ancient Greek, Modern Greek (as well as Medieval Greek before it) has no final consonants in native Greek words other than [n] and [s] (and in high-style usage, [r] too); loanwords can and do have other final consonants, including voiced obstruents, e.g., κλαμπ '(night)club,' κλομπ '(billy) club,' μπριτζ '(contract) bridge,' and the same is true with acronyms (e.g., ΚΕΔ 'Center of Education for Signals' (from Greek Κέντρο Εκπαίδευσης Διαβιβάσεων). Thus as far as Greek is concerned regarding final devoicing, mostly

there is nothing to say or at least very little historical evidence bearing on the topic. However, Newton 1972: 103 notes that there is one dialect of Greek for which final devoicing has been reported, namely that of Kırklareli (Grk Saránta Ekklisies, Blg Lozengrad) in Eastern Thrace, an area where historically there were Bulgarian speakers.[165] In that dialect, according to Psaltes 1905, newly created final stops, arising from high-vowel deletion, underwent final devoicing, as in [fek'] 'shines' from earlier φέγγει. Other northern dialects, however, tolerate final voiced obstruents created by high-vowel deletion; Newton 1972 cites several such forms: [poð] 'foot' (p. 102), [ðod'] 'tooth,' and [šmað] 'sign' (p. 183), for earlier (and StModGrk) πόδι, δόντι, and σημάδι, respectively.

These facts can be interpreted as indicating that final devoicing is an innovation within East South Slavic (Bulgarian and Macedonian) that has spread in a band across Bulgarian- and Macedonian-speaking territory, into parts of Torlak BCMS and west into southern Montenegrin (Greenberg 2000) and adjacent or co-territorial Northwest Geg. It seems to not have reached coastal and Northeast Geg, and parts of Labëri and Çamëri. Its occurrence in the Romani of Macedonia, to a greater extent in West Rumelian Turkish than elsewhere in Turkish, in Meglenoromanian in Greece and North Macedonia, and even in a limited manner in the Greek of Eastern Thrace is consistent with this view of the geographic spread of this phonological development.

From a comparative standpoint, it can be noted that final devoicing is also found in all of West Slavic and most of East Slavic – specifically in Russian and Belarusian but not in Ukrainian. In all of the Slavic languages, however, the occurrence of consonants word-finally resulted innovatively from the loss of word-final short high vowels (jers). Here Old Church Slavonic represents the Late Common Slavic situation in its preservation of jers and therefore its lack of consonants in word-final position. The absence of final devoicing in Ukrainian and Slovene and most of BCMS suggests that the North Slavic and South Slavic phenomena arose independently. Moreover, it is found in modern German and in (ancient) Gothic, also presumably independently, although those languages could be invoked as contact sources for West and East Slavic, respectively.

Thus while final devoicing in Balkan Slavic is part of a widespread Slavic phenomenon, its occurrence in the Romani of North Macedonia and the Greek of Kırklareli (Grk Saránta Ekklisies, Blg Lozengrad) as well as in the dental fricative of WRT is clearly areally conditioned. Its distribution within Albanian and its occurrence in South Montenegrin BCMS are more difficult to characterize. The absence of the phenomenon in the intervening Prizren and Kosovo dialects means that the Montenegrin phenomenon must be viewed either as the remnant of a Slavic

165 Thomason & Kaufman 1988: 218 note that Dawkins 1916: 203 finds final devoicing in Asia Minor Greek, which he (quite plausibly) attributes to Turkish influence. They feel though that "there is some doubt about the validity of Dawkins' claim here, since the change is not unique to Asia Minor Greek"; they refer to an earlier work, Thomason & Kaufman 1976:178, note 5, where they cite the Saránta Ekklisies phenomenon (from Psaltes, via Newton) by way of casting doubt on Dawkins's claim of contact-induced devoicing in Asia Minor Greek. Since contact may have caused final devoicing in Saránta Ekklisies, it is not as compelling a counter to Dawkins as it might be.

dialect continuum that formerly stretched across northern Albania (as hypothesized by Božidar Vidoeski, p.c.), which could then also be the source of northern Geg final devoicing, and it might also be argued that Tosk devoicing is a result of shifted speech habits. This hypothesis is supported by examining the maps of the percentages of Slavic loanwords and toponyms in Albanian dialects in Ylli 1997–2000 and the Slavic toponyms mapped in Seliščev 1931. What emerges from these maps is that the intensity of Slavic contacts with Albanian as evidenced by percentages of loanwords and toponyms corresponds closely to the distribution of final devoicing in Albanian as represented in Gjinari 2007: 112. Of particular significance here is the fact that both final devoicing and Slavic loanwords and toponyms tend not to occur on the coast, where there was resistance to Slavic in general. Thus, final devoicing in the Albanian–Montenegrin highlands as well as in parts of Tosk territory is most likely involved in language contact, and final devoicing in general is the kind of localized Balkanism identified by Friedman 2011a as part of Balkan phonologies as opposed to Balkan phonology.

Although final devoicing is typologically widespread, being a natural type of phenomenon (see Hock 1976: 211) nevertheless its distribution in the Balkans argues for it being the result of contact.

5.4.6 Some Final Thoughts on Segments and the Balkans

It can be argued that Sawicka, for all the value of her compendium of material on comparative Balkan phonology and while certainly correct to focus on details of dialects rather than standard languages, has such a mass of detail that a general picture does not always emerge. In the welter of local dialectal developments, it is less clear that there are any commonalities between different languages taken as a whole, whence our view that there are "Balkan phonologies not Balkan phonology." Yet Sawicka presents the parallelisms and other facts as commonalities (even if on the local level). Focusing on commonalities seems almost to be forcing the issue at times, e.g., discussing loss of an *r*/*rr* distinction in some Albanian dialects without tying it to co-territorial languages, as with Arvanitika losing it due to bilingualism in Greek, even though as far as the Balkans are concerned, it is almost a unique contrast (found otherwise in some Romani dialects of the Balkans and Judezmo); also with regard to [r], it is fair to ask if the loss of this sound in some Greek island dialects (e.g., Samothráki) is really in any way connected with other liquid developments elsewhere in the Balkans. It is hard to imagine that they could be. Thus every putative parallel needs to be considered on its own terms, as we have tried to do here, and judged accordingly.

Further, when evaluating putative segmental parallels, another issue to keep in mind is that of separating the phonological from the phonetic. Banfi 1985: 51, citing Georgiev 1977, has stated that "*Georgiev . . . insiste sulla identità dell'articolazione consonantica nelle diverse lingue balcaniche*" ('Georgiev insists on the identity of consonantal articulation in the various Balkan languages'). Reference to "identity of articulation" suggests that one should look at the purely phonetic

dimension to Balkan segments, even though much of the discussion of the sound structure of the Balkan languages has taken phonetics more or less for granted and worked with distinctive, i.e., phonological, units. There are some instrumental studies on the phonetics of individual Balkan languages,[166] and a few have taken a comparative Balkan approach (e.g., Loukina 2008). Yet, phonetics clearly is important, and one study at least suggests that identity in articulation may be a mirage, even for units that seem to be phonologically identical and which are generally described and reported on in the literature using identical symbols from the International Phonetic Alphabet (IPA).

In particular, for all the fact that affricates identified with the same IPA symbols figure in Balkan phonologies in a number of ways (see for instance §§5.4.2, 5.4.2.1, 5.4.4.2, 5.4.4.7, among others, though we have in some instances used a Slavistically based notation ourselves) and show some apparent commonality across the languages, Joseph & Tserdanelis 2006, in a pilot comparative instrumental phonetic study of the acoustic properties of Balkan affricates, found that from a phonetic standpoint, there were at least three types of affricate-like sounds. To be differentiated are the Greek type, with a long stop portion and a brief sibilant release (for which the symbolization [t^s] was suggested), the Romanian type, with a brief stop portion and a long sibilant release (for which [^{t}s] would be appropriate), and the Albanian type, with the stop and sibilant portions being of nearly equal duration (for which [ts] would be appropriate, perhaps really as two segments rather than one complex segment). It is possible too that Bulgarian offered a fourth type (nearly equal duration but each part rather brief, for which [t͡s] would be appropriate), but only one speaker could be measured. More recordings of more speakers and more measurements are needed, to be sure.

Still, such preliminary results make it seem that if one hugs the phonetic ground too closely, there might be nothing in common involving sound across the languages of the Balkans, even though reference is made repeatedly to similar sounds and even though similar symbols are used in discussions of the sounds. If the affricates can differ as to their acoustic details, the same can be true of the stops, even the /p t k/ (etc.) that Feuillet and Afendras, for instance, noted as being shared among the languages (see above, Example 5.10c). However, that is too extreme a view, since the abstraction involved in looking at phonology and units like the distinctive segments, as opposed to phonetics, does have some correspondence with speaker perception, not to mention with the trained ears of descriptivists as well. Nonetheless, such results show that the physical properties of sounds cannot

166 One can cite, for instance, Tucker 2003 and Chitoran 2002b on Romanian, Dodi 1970 and Kolgjini 2004 on Albanian, Minissi et al. 1982 and Sawicka et al. 2021, 2022 on Macedonian, and Pettersson & Wood 1987abc, Wood & Pettersson 1988, Wood 1996ab, 1998, and Zhobov 2004 on Bulgarian. Arvaniti 2007 offers a comprehensive survey of the literature up to then on Greek phonetics, and there has been much work since then which we cannot fully take note of; see Mary Baltazani's VOCALECT project (http://www.vocalect.eu/?lang=en) for some important examples. Lehiste & Ivić 1986 summarize years of work done by the authors on the phonetics of Serbo-Croatian, mostly on accent and prosody.

be ignored, and that distinguishing between phonetics and phonology in talking about the Balkans, in the final analysis, is crucial to a full understanding of the role that sound plays in the sprachbund.

5.5 Prosody

The term 'prosody' in its broadest sense covers all nonsegmental phonological phenomena in language. Traditionally, that would encompass mainly accent (both pitch and stress), tone, and intonation, i.e., "suprasegmental" aspects of pronunciation, and these constructs certainly play an important, if somewhat understudied, role amongst the Balkan languages. The goal of this section is to present and discuss some relevant facts and findings in this area. However, it must be noted that to the extent that such phonological features as nasalization and length, which do show gain and loss and change under conditions of contact in the Balkans, are to be considered nonsegmental, then they too, *sensu stricto*, are part of a full account of Balkan prosody, even if the decision was made (largely for organizational reasons) to discuss them in the sections on segmental phonology (especially §§5.4.1.1, 5.4.1.4, 5.4.1.6, and 5.4.5.1). Moreover, stress was a relevant factor in the development or the formulation of various segmental features, including the occurrence of schwa (§§5.4.1.1v, 5.4.1.3, and 5.4.1.6), some vowel alternations (§§5.4.3.7 and 5.4.3.9), initial vowel loss (§5.4.3.3), and vowel "reduction" (§§5.4.1.3, 5.4.1.5, and 5.4.3.9).

Still, as noted, there are also developments in the Balkans involving prosody as understood in the more traditional sense, even though there is not the same rich literature to draw on as with segmental material.[167] Many studies have focused on well-known paradigms of what is relevant to Balkan linguistics, looking at possible contact-induced changes involving prosody, parallels among Balkan languages at the prosodic level, and the like. Since prosodic phenomena like stress and intonation can serve as boundary markers, they can be very useful in any sort of communicative act but especially where one participant may have a less-than-perfect command of the language being used. Thus contact-induced changes in prosody are to be expected; indeed, Matras 2009: 231–232 considers prosody to be perhaps the single most contact-susceptible of phonological elements.

There are a few ways, though, in which prosody has come up in studies on Balkan languages that go beyond the description and assessment of potential Balkanisms. For instance, prosody has been taken to underlie, in a causal way, the development of a number of Balkanisms: Klagstadt 1963 made the interesting observation that many of the commonly cited grammatical Balkanisms have a prosodic dimension to them, involving short forms that are prosodically deficient, e.g., being accentless or being unable to stand alone, in ways that

167 Note, for instance, that Sawicka 1997, clearly an indispensable compendium of information on Balkan phonologies, devotes five pages to suprasegmentals as opposed to forty-seven pages to segmental material.

would lead them to be labelled as clitics (enclitic or proclitic) by most analysts.[168] The future tense, for example, is expressed by a prosodically weak and reduced form of a verb of volition (see §6.2.4.1), the analytic comparative utilizes an affix-like unaccented marker (see §6.1.5), object reduplication operates with weak pronouns "doubling" (cross-referencing) full noun phrase objects (see §7.5.1), and even the loss of the infinitive involves replacement by a verb form marked by a prosodically dependent element: Greek να, Slavic *da*, Albanian *të*, Romani *te,* Romance *să/s'/si* (see §§7.7.2.1 and especially 7.7.2.1.3). Similarly, Reichenkron 1962 devotes considerable attention to what he calls the "*Redetaktkurv*," translatable as 'speech-rhythm-contour,' and the "*Redetaktrhythmus*," more literally 'speech.beat.rhythm,' of various Balkan languages,[169] intonationally defined constructs in a language as a general feature, and of a particular type (generally rising or generally falling) that he sees as playing a key role in Balkan convergences. While the prosodic properties of the phenomena discussed by Klagstadt are treated here, their morphosyntactic properties are the focus of Chapters 6 and 7, in the sections noted above and elsewhere. Nonetheless, the sprachbund as originally recognized by Trubetzkoy 1923, 1930 cannot be reduced entirely to prosody, as not all features involve Klagstadtian elements (see also §8.1, footnote 5).

Even so, a primary consideration in examining the Balkan languages has to lie in uncovering and evaluating contact-induced convergences and divergences that the languages show. Accordingly, this section offers an overview of some of the main ways in which prosodic change figures in those accounts.

5.5.1 Prosodic Loss

Most of the cases of prosodic loss in the Balkans have already been treated in the sections mentioned above on nasalization and length. But there are languages and dialects in which once-present tonal distinctions have been lost as well.

In Balkan Slavic, for instance, we start with the reconstruction of the Common Slavic phonological system as having what Schenker 1993: 72 calls, in accordance with the long Slavistic tradition, "phonemic distinctions in pitch (intonation)," involving length (short versus long) and pitch contour (rising, also called "acute," and nonrising, also called "falling" or "circumflex"). Distinctions of

168 We find the term "clitic" to be descriptively useful, though we note that there are theoretical viewpoints (e.g., Zwicky 1994) that deny the need to recognize clitics as a primitive theoretical notion, viewing them instead as simply atypical words or atypical affixes, depending on their properties (see Joseph 2001c, 2002ade for the application of this line of reasoning to facts from Greek that have parallels elsewhere in the Balkans, as explored in Sims & Joseph 2019).

169 While *rhythmus* is straightforwardly 'rhythm,' *takt* has more to do with the placement of the beats within the rhythm, e.g., 9/8 *rhythmus* can have a *takt* of 3-2–2-2 or 2-2–2-3 or 2-3–2-2, etc. Even a simple 4/4 can be realized as, e.g., 2-1–1, 1-1–2, etc. Moreover, the stress within the beats can vary, e.g., the Edirne Romani rhythm *gordel* which is 2-2–2-3 but has stress specifically on beats 1, 5, and 7 (Seeman 2019: 325), i.e., **1**-2 3-4, **5**-6, **7**-8–9.

both tone and length are continued in much of Slovene and the northern half of BCMS, although Slovene also has a nontonemic system (Priestly 1993: 389ff.) and Zagreb and Belgrade colloquially also lack tone (Browne 1993: 382ff.). While the inherited tonal distinctions of late Common Slavic were lost in most of BCMS, a new tone distinction from retracted stress (the so-called neo-Štokavian acute) spread over most of BCMS territory. However, peripheral dialects to the north and along the coast (except parts of Istria) retained older tonal distinctions, while those to the south (Southern Montenegro, the Sandžak, northern Kosovo, and eastern Serbia along the east bank of the Ibar northward as far as southeastern Vojvodina – as well as western Istria) lost tone but kept length and stress (for details see Ivić 1958, 1985). It is precisely in the Torlak (Prizren-Timok) dialects of BCMS, as in Bulgarian and Macedonian, that both tone and length were lost. This loss of distinctive vowel length is a key feature linking Torlak BCMS with the rest of Balkan Slavic. It was a major reason that Ivić 1958, which is limited to Štokavian BCMS, does not include Torlak (although Ivić 1985 does, under the label Prizren-Timok, as did the first, 1956 edition). The decision to exclude Torlak from the main body of Štokavian was motivated by its heavy Balkanization, but especially by this loss of phonemic length, which is otherwise found only in Macedonian and Bulgarian of the South Slavic languages (cf. Friedman 2006d and the references therein, also Afendras 1968: 100). However, as Vidoeski 1999b: 117–124 notes, new phonemic length has arisen in some Macedonian dialects owing to various types of later elisions. These facts do not, however, change the shared innovation of elimination of length in Balkan Slavic, including Torlak BCMS.

While shared loss is not diagnostic in the same way as an additive shared innovation, in the context of South Slavic the loss of both tone and length distinctions looks particularly Balkan. The fact that these same losses occurred in East Slavic and (later) in Lekhitic is beside the point, since both Czech and Slovak and Slovene and northern BCMS remained relatively conservative, and Slovene and northern BCMS most conservative of all. The loss of Common Slavic tone and length in Balkan Slavic must therefore be seen as an independent innovation, and one that occurred at the Balkan peripheries of Slavic territory.

While the loss of a pitch-based accent and its replacement by an intensity-based stress occurred relatively early in Greek – the changes are usually dated to the Koine period (Horrocks 2010: 117–118), thus predating the period of sprachbund-related contact by several centuries – the facts from Greek prosodic change might be taken to suggest that the Balkan Slavic shifts owe something to Greek influence. This is all the more striking when we take into account the limitation of stress to the final three syllables, which is characteristic in varying realizations of the entire southwestern region of Balkan Slavic, i.e., Macedonian. Another view of the relevance of the Greek facts, especially when the rest of Slavic is taken into account, is that they show that prosodic systems can change on their own, without there necessarily being language contact as an impetus. Still, the basic observation is that there have been changes in the prosodic systems of several of the Balkan languages, and while some of them might have been independent of one another, it

is undeniable that there is a regional clustering of loss of both length and tone in the Balkans and the fixing of stress no further back than the antepenult (with minor exceptions) as opposed to length and freer or even fixed initial stress north of the peninsula. The loss of Latin length in Balkan Romance and also of Albanian length in much of Tosk can also be mentioned here. In the case of Romani, Indic length was lost before arrival in the Balkans, but Romani north of the Balkans in contact with languages with distinctive length have developed new length as a result of contact (Matras 2002: 59–60).

5.5.2 Accentual Realization, Mobility, and Adjustments in Domain Extensions

As suggested in the previous section, the realization of accent now across the Balkan languages is an intensity-based, stress accent. Feuillet 1986: 50–51, in his list of phonological Balkanisms, explicitly refers to this as "*accent dit d'intensité*" ('accent said [to be] of intensity') and notes that it is "*partout de type dynamique*" ('everywhere of a dynamic type'). He does recognize that this accent "*comporte aussi des éléments musicaux*" ('includes also musical elements'), i.e., pitch and tonal dimensions, but he is quick to point out that these "musical" aspects do not have a distinctive value.[170]

Important also, and of wide distribution in the Balkans, is the fact that the lexical realization of this accent involves mobility, that is, variable placement in related words, typically conditioned by the lexical specifications or the grammar in ways that are not purely phonological in nature. Standard Macedonian, and, more importantly, the western dialects on which it is based, are usually labeled as an exception to this claim about mobility, in that it has fixed antepenultimate stress; as becomes clear below, though, related word forms in (western) Macedonian do show a certain type of accentual mobility, just not one that is grammatically determined – rather it has a phonological basis (see (5.12) below). Moreover, with the exception of Bulgarian, the positioning of the accent is restricted as to which syllable within a word it can fall on, and specifically limited to occurring on one of the final three syllables.[171]

These aspects of Balkan accent can be illustrated with examples from the various languages. In Greek, for instance, accent shift (mobility) is found in derivation, as in the noun όνομα 'name,' with antepenultimate stress in this nominative/accusative form, as opposed to the related verb ονομάζω 'I name,' with penultimate stress in the present tense, or in inflection, as with the nominative όνομα vs. genitive singular ονόματος / genitive plural ονομάτων /, or present tense ονομάζω vs.

170 There is some instrumental research into accent in some of the languages that provides a basis for talking about the acoustic cues for this dynamic stress of intensity. For instance, Arvaniti 2000: 9 has argued for Greek that the "amplitude integral, a measurement that combines those of duration and (average) amplitude" is the main acoustic correlate of stress (see also Arvaniti 2007: §4.1), and Lehiste & Ivić 1986: passim discuss the role of intensity in standard Serbo-Croatian.

171 There are also occasional, and even systematic, exceptions in some of the other languages, as discussed below.

imperfective past ονόμαζα 'I was naming.' The governing factor is a combination of the number of syllables in the word and the particular suffixes involved or the grammatical categories the word expresses. That is, in Greek, the accent can fall only on one of the last three syllables in a word (the modern reflection of the Ancient Greek *Dreimorengesetz* 'three-mora rule' or Law of Limitation) and while the basic placement (e.g., as shows up in the nominative singular) in the stem is marked and (presumably) stored in the lexicon, deviations away from that basic placement are governed by affix-type and grammatical information. Thus the -ά- of the denominal verbalizing suffix -άζ- is always accented in the present tense (so, 1SG ονομάζω, 2SG ονομάζεις, 1PL ονομάζουμε) but that accent placement generalization is overriden in the past tense which, for this type of imperfective past, has antepenultimate accent, as in ονόμαζα 'I was naming'; the accent is only accidentally back on the -α- of the derivational suffix in the past plural, since the plural endings add a syllable and thus occasion – due to the modern Law of Limitation – a shift relative to the singular, to, e.g., ονομάζαμε 'we were naming.' The shift between όνομα and ονομάζω or between όνομα and ονόματος, or for that matter seen in όνομα / ονόματος / ονομάτων, is triggered by the particular suffix involved whether derivational (as with -αζ-) or inflectional, but again within the limits of the final three syllables.

This type of conditioning on accent placement also means that accent can be distinctive, with different word forms distinguished by the positioning of the accent. This effect can be lexical, as in νόμος 'law' versus νομός 'prefecture,' or grammatical, as in αγαπά 'love.PRS.3SG' versus αγάπα 'love.IMPV.SG,' but in either case, accent placement is distinctive.

Similar facts, *mutatis mutandis*, are to be found in the other languages, for the most part. In Albanian, for instance, accent shifts occur only between derivationally related words (see also §5.2),[172] as in *habí* 'surprise' vs. *habitór-* 'admirative' or *katúnd* 'village.NOM.SG'versus *katundár* 'peasant.M,' *katundarí* 'peasantry,' with the stress generally falling on the final syllable of the stem, but *katundáre* 'peasant.F.' When inflectional endings are added, there is no effect on the placement of accent, and it remains on the stem-final syllable, so that the genitive plurals of these words are *katúndeve*, *katundárëve*, *katundaríve, katundáreve*, respectively. Moreover, stress placement is distinctive, as in the pairs *pára* 'before'/*pará* 'money,' *béla* 'spades'/*belá* 'trouble,' *xhakoní* 'seminary studies'/*xhakóni* 'the seminarian.' In some instances, in such pairs, the stem-final-syllable-placement rule is observed: in *xhakoní*, *-i* is the stem-final element, whereas in *xhakóni*, the stem is *xhakón-* and the *-i* is the nominative definite article morpheme; in other instances, loanwords have created the minimal pair, in that in *béla*, the stem is *bél-* (nominative singular *bel*) and *belá*, a borrowing from Turkish *belâ* (with final stress), then comes to contrast with 'spades.' There are what amount to lexical exceptions to stem-final stress: in the adverbial *pára*, there is no clear morphemic

172 Accent is indicated throughout for expository purposes, though it is not a part of the standard orthographies for the languages involved.

division so that the stem is disyllabic *para-* (the noun *pará* is a Turkish loan) and the manner adverbial suffix *-azi* (see footnote 8) does not attract stress off of the adjectival base it attaches to; so, too, the feminizing suffix illustrated above.

Romanian shows a similar situation, matching Albanian rather closely, to the extent that Hamp 1989a: 47, as noted above in §5.2 (see especially (5.2) and (5.3)), claims that "historically Romanian is Latin spoken with an Albanian stress system." Under this interpretation, Albanian and Romanian accentuation would be a contact-induced prosodic parallel in the Balkans, admittedly at a rather ancient time-depth, as some part of an ancient Albanoid population shifted to Latin to produce Romanian (and the rest of Balkan Romance). With regard to present-day Romanian, an example of a minimal pair showing the distinctiveness of accent placement, as well as the grammatical basis to the positioning of accent, is present *súnă* 'it sounds' vs. past *sunắ* 'it sounded.'

As for Bulgarian, the description of stress given by Scatton 1993: 193 is that "word stress is dynamic: stressed syllables are louder and longer and have a higher fundamental frequency than unstressed syllables." Moreover, he notes that stress is "free" as to its placement in a word; examples like *cárevica* 'corn' show the lack of a restriction like that found in Greek or, importantly, (western) Macedonian. Bulgarian stress is mobile too, shifting between different syllables in related words, generally with a grammatical basis. A minimal pair from Bulgarian, making the same point about distinctiveness and grammatical determination as the Romanian examples, is present *četé* '(s)he reads' vs. past *čéte* '(s)he (did) read.'

The situation is somewhat different in Macedonian, as noted above, with what is usually described as fixed antepenultimate stress in the standard language (Koneski 1967: 141). Thus, with certain exceptions, the accent is always on the third syllable from the end in polysyllabic words, and a comparison of derivationally and inflectionally related forms shows this clearly: *vodéničar* 'miller' but with the plural suffix *vodeníčari* 'millers' and with a definite article *vodeničárite* 'the millers' (if we take the view that the article is a suffix; see §6.1.2.2.1). There are lexical exceptions to this generalization, as Friedman 1993d emphasizes, generally in loanwords, and these have led to some minimal pairs, such as *kraváta* 'necktie' vs. *krávata* 'the cow' (morphologically, *kráva* with the definite article *-ta*), but generally, especially in native vocabulary, the antepenultimate rule holds.[173] These accentual patterns for Macedonian, then, are rather like what happens in certain past tense forms in Greek, as given above, where an antepenultimate rule is in effect giving 1SG ονόμαζα but 1PL ονομάζαμε; Macedonian differs from Greek in that the antepenultimate rule holds over virtually all lexemes and grammatical categories, not just a few as in Greek. These points of comparison and contrast between Macedonian, Bulgarian, and Greek, with some

173 A native exception to this stress generalization is the verbal adverb and some other adverbial expressions; see §5.4.1.2 above and especially footnote 35. In some dialects, e.g., Tetovo, however, the verbal adverb is regularized.

relevant data from Balkan Romance too, are of interest typologically within the Balkans, even if there is not necessarily any historical or contact-related connection among them.

As for Turkish, however, although the precise analysis of stress in the standard language is complicated (see Göksel & Kerslake 2005: 26–34 on this), it is generally perceived in the Balkans as having stress on the word-final syllable, as the evidence of stress in loanwords shows, e.g., Turkish *kebap* (ACC *kebabı*) 'roasted meat, etc.,' gives Alb *qebáp,* Aro *cibápi,* Blg *kebáp,* Grk κεμπά(μ)π, which all refer to some sort of grilled meat. In the case of some varieties of Balkan Turkish, however, the situation is different. Tufan 2007: 92–93 shows that the stress in Gostivar Turkish (western North Macedonia) follows the same ante-penultimate stress rule as the Macedonian with which it is in contact, with the same sort of mobility, e.g., WRT *gélırler* 'they come,' *gelírsınız* 'y'all come,' *kâşikçi* 'spoon-maker,' *arábaci* 'mechanic [for cars]'; the same is true in Ohrid-Prespa (Ahmed 2012). Thus these WRT dialects with antepenultimate stress have clearly been influenced by western Macedonian speech patterns.

What is especially interesting in western Macedonian are the developments in what can be termed extended phonological domains, when phrasal units condition accent shift. The examples of plurals like *vodeníčari* 'millers' above show that the accent moves as the phonological domain expands; as noted in footnote 37 above, we refer to this movement, and other related accentual developments in these extended domains, as accentual adjustments. According to the prescribed standard, these adjustments occur even with the addition of elements that are not affixes – though in some analyses some of them might be – or which appear to be prosodically deficient words (often called *clitics*; see footnote 168), as well as across boundaries of full words in noun phrases. In particular,[174] if the definite article is considered a clitic rather than a suffix, then it is an example of an accent shift due to a clitic, as in (5.12), which would accord with prescribed accent shifts to clitics with various verb forms extended by weak object pronouns attaching to their right or left, as in (5.13abc),[175] or with indicative verbs extended by the negation marker attaching to the left of a less-than-trisyllabic verb, as in (5.13def), with interrogatives (5.13gh), prepositions (5.13i), presentationals (5.13j), and noun phrases (5.13kl):

(5.12) a. vodeníčar-ot 'the miller' (cf. *vodéničar* 'miller')
b. vodeničári-te 'the millers' (cf. *vodeníčari* 'millers')

(5.13) a. dónesi 'bring!'
b. donési_go 'bring it!'
c. donesí_mi_go 'bring me it' (but also *donési_mi_go* by analogy with b)

174 In these and later examples, where useful for expository purposes, the domain-extending morpheme (article, pronoun, etc.) that provokes the accentual adjustment is set off with a hyphen or underscore, contrary to standard orthography.

175 As Friedman 1993d: 254 notes, the verbal adverb, which also takes weak object pronouns on its right, does not show the same accentual adjustment as the imperative, e.g., *noséjkji mu go* 'while-carrying to-him it.'

d. sákam ‘I want’
e. né_sakam ‘I don’t want’
f. ne_mu_já_dade ‘s/he did not give it to him’ (also *ne_mu_já_dal* for the M.LF)
g. štó_sakaš ‘what do you want?’ (also *što_mú_reče* ‘what did you say to him?’)
h. kolkú_pari sakaš ‘How much money do you want?’
i. só_nego ‘with him’(“obligatory”) *megjú_nas* ‘between us’ (“facultative”)
j. evé_ti_to ‘here he is [for you]’
k. celó_leto ‘all summer’
l. kiseló_mleko ‘yoghurt’ vs. kíselo mléko ‘sour milk’

We can note here that for most of these accentual units, the prescribed norm is now a regionalism or even a dialectism, and, while taught in schools, it is mostly not used by educated speakers in practice. According to Petroska 1998: 64, accentual units are only preserved with interrogation and negation, while Sawicka et al. 2021: 95 only found them under negation, and even then not consistently. Current educated usage argues against treating articles as clitics in this respect, since prescribed usages such as *novatá_kukja* ‘the new house’ never occur in educated speech, but rather *nóvata kúkja*. The original prescriptions are given in Koneski 1967: 139–210, and Jovanova-Grujovska & Vojneska 2017: 95–96 base the current prescriptions on Koneski 1967. The situation in western Macedonian contrasts sharply with Standard Bulgarian (based on its northeastern dialects), which does not alter the stress placement in such extended domains, except for the negative particle, which can attract stress onto a clitic, e.g., *Áz ne gó vidjáh* ‘I did not see him’ (lit., ‘I not him saw’), *Áz ne šté mu go káža* ‘I will not tell him it’ (lit., ‘I not will him it tell’), etc. (Hauge 1999: 194–195). In this sense Bulgarian accent placement is relatively free, whereas in Macedonian it is subject to one form or another of the antepenultimate restriction in the West. Eastern Macedonia is transitional between the two systems. Stress is mobile, but limited to the last three syllables (or penult and antepenult in some environments in the western part of that region) and is paradigmatically fixed in the western part of that region. The extreme southwest has fixed penultimate rather than antepenultimate stress. One way of analyzing this difference is in terms of what counts as a phonological word in the standard languages, which represent in this respect three dialectal extremes.

Romanian offers a somewhat limited set of data showing accent shift occasioned by the extension of the prosodic domain, in the manner of Macedonian. The extension of the domain occurs with the addition of a weak object pronoun and the 2PL is especially susceptible (Lombard & Gâdei 1981: II.97 n.), so that the reflexive verb *a me teme* ‘be afraid’ has a 2PL imperative *teméţi-vă – vă* is the domain-extending reflexive pronoun – where the verb form by itself would be stressed on the leftmost vowel, thus *témeţi*. Aromanian, like Greek, is subject to the three-syllable rule, i.e., the stress is free but cannot move further back than the antepenult (Gołąb 1984a: 46).

Greek also shows accentual adjustment in extended domains, thus producing a superficial resemblance to the context for Macedonian accent shift, but differing in the details of how the domain extension affects accent. In the Greek case, the extension to the domain typically comes with the addition of genitive case

possessive pronouns after a noun, as in (5.14), or the attachment of weak object pronouns after an imperatival verb or a verbal adverb, as in (5.15):

(5.14) όνομα ‘name’ / όνομά του ‘his name’
his

(5.15) a. βλέπετε ‘See!’ / βλέπετέ τον ‘See him!’
IMPV.PL IMPV.PL him
b. βλέποντας ‘while-seeing’ / βλέποντάς τον ‘while seeing him’
VBL.ADV him

Some dialects, moreover, such as Cretan but also more northerly dialects such as Thessalian (Newton 1972: 198–199), have the accent adjustment in longer inflected forms, where the relevant domain is extended by a disyllabic person/number affix, as in (5.16):

(5.16) έρχουμάστι ‘we come’ (cf. StModGrk ερχόμαστε)
1PL

There are aspects of these patterns that are somewhat controversial, especially concerning the nature of the leftmost accent in Greek, e.g., whether it is reduced phonetically and whether its realization coincides with a putative rhythmic accent supposedly added every other syllable in longer words, but experimental work by Arvaniti 1992 (see also Arvaniti 2007: §4.3) confirms that the leftmost accent is weaker than its counterpart in the unextended domain, and moreover that it is distinct from any rhythmic accent that Greek may have (in fact, Arvaniti argues from her findings that there is no rhythmic accent rule for Greek). Thus, a more accurate representation of the forms in (5.14)–(5.16) would be ὸνομά του (etc.). Importantly from the Balkan perspective, these Greek patterns differ from Macedonian as to which categories are affected, in that Greek includes the verbal adverb as a “host” for domain extension but Macedonian does not, and as to the effect itself, in that Greek adds a primary accent while Macedonian shifts the place of the accent. Although the realization itself of accent as stress is a Postclassical innovation in Greek, as noted in §5.5.1, the accentual adjustment in extended domains is an old feature of Greek, being the modern reflection (with some alterations) of accentual adjustments with enclitics found in Ancient Greek.

Based on these facts, therefore, despite some similarities, the Greek and the Macedonian developments would seem to have to be taken to be independent of one another. When one looks to dialects of Macedonian, however, as well as some of Bulgarian, a slightly different picture emerges that is suggestive of Greek influence. A closer examination, though, nonetheless reveals a more likely judgment in favor of the independence of the phenomena.

That is, there are Balkan Slavic dialects that show a double accent in extended domains, superficially rather like the Greek phenomenon. This has been discussed in the literature for over a century but has been reexamined and taken up most

recently in a series of important studies by Alexander (e.g., 1999, 2002). Alexander 2002 distinguishes between "double accent" (DA), found with domains extended on the right, as in (5.17) from the Macedonian dialect of Bansko, in the Republic of Bulgaria, and two kinds of "additional accent," one found quite widely (even prescribed for literary Bulgarian) with domains extended on the left, as in (5.18), and one found in the eastern Bulgarian dialect of Erkeč almost always with the definite article on the right and with that article accented, as in (5.19):

(5.17) víkamé go 'We call him'
call.1PL him

(5.18) ne gó poznávam 'I don't know him'
NEG him know.1SG

(5.19) dóktori-té 'the doctors'
doctor.PL-DEF.PL

The type of (5.17), what Alexander 2002 calls "canonical DA," matches the Greek pattern exactly, as a comparison with (5.15) shows. Moreover, in Bansko, one finds other realizations of DA that match the northern dialectal Greek type of (5.16), involving two accents on a single long word, as in (5.20):

(5.20) krástavíca 'cucumber'

However, Bansko also has DA that is quite unlike Greek, involving two accents on a single shorter word (5.21a), more than two accents across an accentual domain (5.21b), accent on the preantepenultimate (5.21b),[176] and DA when the head form is disyllabic and the extended domain only trisyllabic (5.21c):

(5.21) a. kázvamé 'we say'
say.1PL.PRS

b. práznuválo sé e 'it was celebrated; there was celebration'
celebrate.LF. N INT R is

c. pásmo-tó 'the skein'
skein.SG-DEF.SG

Moreover, the two types of "additional accent," illustrated above in (5.18) and (5.19), are not paralleled in Greek. All of these differences make it hard to consider the Greek and the Balkan Slavic phenomena as being the same, and they thus make a contact-based explanation, with Greek influencing the Slavic, less compelling overall.[177] And, indeed, this is the line of argumentation that

176 Some northern Greek dialects, including Crimean Greek, and others too, e.g., Rhodian and the Greek of southern Albania, show accent farther back in the word or extended domain (i.e., farther from word-end or domain-end) than the antepenultimate; see Newton 1972, Delopoulos 1983, and Joseph 2001b for examples and discussion.

177 So also, of course, for considering these facts to show Slavic influence on Greek; since the accentual adjustments have a historical antecedent in Ancient Greek; however, most assumptions of a contact-based explanation have looked to Greek as the source for Slavic, not vice-versa.

Alexander takes, quite reasonably, though Sawicka 1997: 74 states confidently that "the Greek influence on Macedonian [in this case] is unquestionable." It perhaps cannot be proven that Greek had any influence over the Slavic developments, but it is a fact that Greek was once widely spoken, well into and throughout the nineteenth century, in the areas of Balkan Slavic where the multiple accent phenomenon is found. Bansko is about forty kilometers from Melnik, for instance, a Greek town before the Greek-Bulgarian population exchanges right after World War I, and even the isolated eastern dialect of Erkeč is near enough to one-time Greek enclaves on the Black Sea coast. We can speculate, then, that Bansko speakers innovatively took the *víkamé go* and *krástavíca* accentual adjustment types – (5.16) and (5.19) – from Greek, with their familiarity with and use of Greek triggering reverse interference onto their native Slavic, but that they then further extended this multiple accent realization in different directions, going beyond anything found in Greek; this would not be at all unusual, for once a feature enters a language from outside, it can indeed go its own way and can diverge from the source feature in the donor language.

Thus, even though inconclusive and perhaps owing to an independent development in each language, the Greek and the Balkan Slavic accentual readjustments, viewed in the larger context of the realization of the mobility that the realizations show, offer an intriguing case of prosodic parallelism in the Balkans, whether or not convergent due to contact.

In the case of some WRT in western North Macedonia, however, there is a clear influence of the western Macedonian accentual pattern on the former, as noted above. To this we can add that accentual units such as those illustrated in (5.13de), have been recorded in Skopje Bugurdži (Bodnárová 2018a). In this dialect, *lazámi* (from Trk *lâzım*) is the word for 'need' and follows the usual stress pattern for Romani dialects of the Balkans (final on pre-Byzantine loanwords, nonfinal on post-Byzantine loans, with inflections of clitic origin not stressed). In a negated sentence, however, we find *Na lazəmí nekoj* 'I don't need someone [to do something for me],' a situation that reflects an antepenultimate pattern in negated sentences, as seen in (5.13e). Here we can note that the place of stress is an example of compartmentalization in the grammar of Romani (Friedman 2013b). Moreover, this seems to be a boundary marker precisely in the Balkans, since stress in Central Romani dialects to the north tends to follow the stress pattern of the major contact language. (cf. Matras 2002: 62–64).

5.5.3 The Prosody of "Clitics"

As noted in §5.5.1, many of the commonly cited morphosyntactic Balkanisms involve small prosodically dependent elements that show quite similar behavior across the various languages. For instance, the definite article is postposed insofar

as it occurs after the first element in the noun phrase where it can occur, as in Example 5.22:[178]

(5.22)
a. Romanian:	om-ul		‘the man’
	man-the		
b. Macedonian:	grad-ot		‘the city’
	city-the		
c. Bulgarian:	pesen-ta		‘the song’
	song-the		
d. Albanian:	kal-i		‘the horse’
	horse-the		

Given these facts, it is indeed fair to describe the article as postposed relative to the noun, and even fuller noun phrases, with modifiers, give that impression; not all of these phrases show the most typical placement for particular languages and some may have marked pragmatic effects of emphasis or the like, but the point is to show the positional possibilities of the article (5.23):

(5.23)
a. Romanian:	om-ul	bun		‘the good man’
	man-the	good		
b. Macedonian:	grad-ot	nov		‘the new city’ (marked order and effect)
	city-the	new		
c. Bulgarian:	pesen-ta	xubava		‘the nice song’ (marked order and effect)
	song-the	nice		
d. Albanian:	kal-i	i	madh	‘the big horse’
	horse-the	PC	big	

It is possible, however, and in some of the languages more usual, for the adjective to occupy initial position in the noun phrase, and in that case, the definite article attaches to the adjective and not to the noun (5.24):

(5.24)
a. Romanian:	bun-ul	om	‘the good man’	/	*bun om-ul
	good-the	man			
b. Macedonian:	nov-iot	grad	‘the new city	/	*nov grad-ot
	new-the	city			
c. Bulgarian:	xubava-ta	pesen	‘the nice song’	/	*xubava pesen-ta
	nice-the	song			
d. Albanian:	i madh-i	kal	‘the big horse’	/	*i madh kal-i
	PC big-the	horse			

What these facts show is that the definite article is positioned relative to the relevant elements making up the noun phrase, not to the noun proper; that is, it is

178 Balkan postposed definite articles are unquestionably of clitic origin. Their current status in the respective Balkan languages (clitic versus affix) is a topic of debate, and depends in part on how these elements are defined theoretically. The ambiguity of their status can be seen in the characterization by Lunt 1952: 41, who writes of the Macedonian definite article forms that “They are enclitics and can be termed suffixes.” Halpern 1992, 1995 has relevant discussion, working with clear definitional criteria. See §6.1.2.2.1 for further relevant discussion of this ambiguous status, as well as §6.1.1 (especially Table 6.2), §6.1.1.2.1.3 (especially Table 6.6), §6.1.1.6, footnote 102, and §6.1.2 (passim), for more on the article. In §7.3.3, there is further brief discussion of terminological (and theoretical) points, where the elements in question are weak object pronouns, and §7.5 has material on clitics in the Balkans more generally. See footnote 168 for some theoretical considerations.

in second position in the noun phrase, occurring after the first inflectable element. This sort of behavior, involving occurrence in second position relative to some domain, is quite familiar cross-linguistically with prosodically deficient and dependent items, and is generally referred to as "Wackernagel" positioning or labelled as an instance of "Wackernagel's Law," after the formulation by Jacob Wackernagel 1892 of the behavior of unaccented forms in ancient Indo-European languages. There are of course other ways in which the prosodic weakness of such forms could be resolved, but Wackernagel second-position is an especially common strategy, even if it is altered via reanalysis in some languages: in Greek, Albanian, Balkan Romance, and western Macedonian, for instance, in the finite verbal complex, initial position placement has replaced second-position strategies.

Similarly, other elements that figure in Balkan morphosyntax are positionally restricted, and thus not freely occurring. They therefore show prosodic dependence and require the support of a host element. Since the presence of accent is often taken as indicative of lexicality,[179] one can take prosodic strength to correlate with a form being a fully viable lexical item in its own right. For instance, the invariant future marker – Albanian *do*, Aromanian *va*, Meglenoromanian *s'/ãs*, Greek θα, Macedonian *kje*, Romani *ka*, and Romanian *o* – must appear to the left of the main verb, and is prosodically dependent on that verb, being unable to occur by itself, for instance in ellipsis.[180] This situation is illustrated in (5.25)–(5.27) with facts from Greek, Macedonian, and Albanian respectively, but for the most part can be exemplified across the other languages as well:[181]

(5.25) a. θα βλέπω 'I will see' / *βλέπω θα
b. θα βλέπεις τον Γιάννη? Ναι, *θα 'Will you see John? Yes, I will [see John]'
FUT see.2SG the.ACC John.ACC yes (cf. Eng *Yes, I'll)

(5.26) a. kje vidite 'you will see' / *vidite kje
b. kje go vidite Ivan? Da, *kje Ø 'Will you see John? Yes, I will [see John]'
FUT him see.2SG John yes (cf. Eng *Yes, I'll)

179 Alexander 2002: 4, for instance, states that "As in most linguistic systems, the presence of a word accent in Balkan Slavic signals the existence of a lexical word." This can be taken as an argument that clitics, insofar as they affect the accent, are words, albeit prosodically deficient (and thus "atypical" – see footnote 168) words.

180 Albanian *do* as a main verb meaning '(s)he wants' should be considered a different item, which can, of course, occur by itself. We can also note here that Miletič 1934 treats the postposing of a conjugated future marker after a short infinitive as relatively normal. At this point, it is completely archaic.

181 Bulgarian dialectally and archaically can have a postposed future marker, and it also occurs in the probabilitive mood of Novo Selo. Apparent postposing in non-Torlak BCMS is rather a matter of the future marker being in second position in a clausal domain. See §6.2.5.8 for some details and discussion of the probabilitive. The matter of the ordering of these elements is taken up as well in §7.4.1.2.2.2, and the future specifically is discussed in §7.4.1.2.2.2.2.

(5.27) a. do (të) shikosh ‘you will see’ / *(të) shikosh do
b. do ta shikosh Gjonin? Po, *do (të) Ø
FUT DMS.him see.2SG John Yes
‘Will you see John? Yes, I will [see John]’ (cf. Eng *Yes, I’ll)

The weak object pronouns also show a similar prosodic dependence, and concomitant nonlexicality, in that they are unable to occur on their own, are unaccented, and always depend on a host. Just like the future marker in (5.25)–(5.27), they cannot, for instance, be elliptical one-word answers in questions, as shown in (5.28a), though strong, accented, forms of the pronouns can function in that way, as in (5.28b); again, the behavior of strong and weak forms in ellipsis is illustrated with one language, here Macedonian, but the pattern of grammaticality and ungrammaticality indicated here is duplicated in all the languages:

(5.28) a. Dali go gledate nego ili ja gledate nea? *Go / *Ja (Mac)
Q him.WK see.2PL him.STR or her.WK see.2PL her.STR him.WK her.WK
‘Do you see him or do you see her? *’Im. / *’Er.’
b. Dali go gledate nego ili ja gledate nea? Négo. / Néa.
Q him.WK see.2PL him.STR or her.WK see.2PL her.STR him.STR her.STR
‘Do you see him or do you see her? Him. / Her.’

However, the weak object pronouns show some mobility in all of the languages, as shown for Greek in (5.29a) and Bulgarian in (5.29b):

(5.29) a. τον βλέπεις / βλέπε τον (Grk)
him.WK see.2SG see.2SG.IMPV him.WK
‘You do see him’ ‘See him!’
b. zova go / Az go zova (Blg)
call.1SG him.WK I.NOM him.WK call.1SG
‘I call him’ ‘I call him’

The conditions for this mobility can be specified in precise structural terms, though varying from language to language. For instance, the positioning is grammatically determined in Greek and Macedonian, based on the type of verb (finite or nonfinite in the formulation of Joseph 1983a: chapters 2, 5), but in Bulgarian, as with the definite article facts above, the positioning of weak object pronouns relative to the verb is more Wackernagelian in nature; in fact, though, whenever a finite verb would be the second element, the rule that weak pronouns must precede it means that they will appear to be in Wackernagelian position.

All of these facts are striking and such prosodic similarities among the Balkan languages have been noted in the literature frequently (e.g., by Joseph 1983a: 238–239; Sawicka 1997: 76, among others). Still, despite the apparent convergence, it is not clear that language contact is responsible for the prosodic side of the behavior of these elements.

Prosodic weakness is an inherited trait with some of these elements, e.g., Proto-Indo-European, and after it Common Slavic (to judge from the evidence

of Old Church Slavonic) and early Greek, at least, had weak object pronouns (the basis for Wackernagel's observations about their placement, along with Sanskrit evidence). It is true that the specific forms found for the pronouns show the effects of change – there was some chronological and author-based (maybe dialectal) variation within Ancient Greek, for example, for the third person singular accusative form, with έ occurring in Homer and somewhat later (Plato), μιν also in Homer and Herodotos, and later still, αὐτόν (м) – whereas the modern form is τον (masculine), την (feminine), το (neuter); see Dressler 1966 and Bubenik 2001 for particulars. Moreover, some aspects of their behavior, and in particular their positioning, have changed – weak object pronouns were positioned in Ancient Greek and early Slavic relative to the clause whereas now, in Balkan Slavic and Modern Greek, they are positioned relative to the verb as the governing domain (cf. Horrocks 2010: 277–281; Pappas 2004a). Still, it seems that there has always been a category of prosodically deficient pronouns in the various languages and their behavior in part simply reflects this inheritance. Similar considerations hold for some of the other elements in question.[182]

Furthermore, the behavior of prosodically weak elements cross-linguistically is quite uniform, in general terms at least, and second-position, which has come to be called Wackernagel's Position, is a particularly common landing site for such elements.[183] Thus, it seems that there is actually a relatively small range of "solutions" for what to do about prosodic weakness, and as a result, the criteria of universality and naturalness must be taken into consideration here. In this case, they make it hard to say for certain that contact is involved in the prosodic convergences seen here.[184]

And certainly from the point of view of the mechanism of contact involved, one has to wonder how these observed convergences would arise in a contact situation. It is conceivable that a "template" with the various elements, say X-article-Y in the case of the postposed definite article (where X or Y could be a noun or an adjective), in

182 For instance, one type of additional (secondary) accent described in Alexander 2002: 5 involves an extended domain to the left, with the negative marker *ne* as the leftmost element. In this case, there are independent historical antecedents for special prosodic behavior of a negation marker, such as the Ancient Greek negative οὐ being generally "proclitic," a use that may underlie the positioning and prosodic weakness of Modern Greek negator δεν that occupies the same slot. More germane, though, is the fact that in Baltic and elsewhere in Slavic, and thus presumably in Proto-Balto-Slavic, the negative marker **ne-* has special prosodic properties; see e.g., Nevis & Joseph 1993 on Lithuanian, Wackernagel affixes, and the ambiguous status of *ne-* in that language.

183 Relevant references include van Riemsdijk 1999 – which has discussions of Balkan languages, though not in a comparative or historical framework – and likewise Halpern & Zwicky 1995 and Steele 1975.

184 Still, the distribution of shift to first position with finite verbs is highly suggestive of language contact (especially when the postposing with verbal adverbs, a characteristic of Greek and Macedonian, is taken into consideration).

a particular order could be implicitly available to borrowing speakers by virtue of their having several exemplars of noun phrases with these elements in this order. But the material in all of the cases discussed (and those not discussed even) is in each instance native material in the particular languages, e.g., for the article: *-ul* for Romanian, *-ot* for Macedonian, *-i* for Albanian, and so on. Thus the borrowing would have to have been at a more abstract level or else some degree of metalinguistic awareness of the borrowers, i.e., some analysis of what they borrowed, would have to be assumed. This of course is exactly what goes on in calquing, as argued in §3.2.1.7, though always based on a concrete, surface model, but one has to ask whether the special prosodic properties derived from this new pattern or if instead native material, already with the relevant prosodic features, was pressed into service. Thus structural borrowing per se does not seem to be a promising avenue for explaining the observed prosody. The situation would be different in a substratum scenario – which, as Hamp 1982 has argued (see also §1.2.1 and §6.5) is in fact the right way to think about the definite article convergence – as in that case speakers would look for material in L2 to plug into their native pattern.

In the case of the future marker, the developments that led to the invariant *do*, *va*, *o, šte*, *kje*, θα, *ka,* etc. are demonstrably language-particular; the stages with fuller forms and various reductions are observable in the textual record for Greek, Balkan Slavic, and Balkan Romance, and there is no reason to think that the history of Albanian or Romani would be any different if the documentation were available (and see §6.2.4.1 for details). In Greek, for instance, where the richest documentary record is available, all of the stages between the fullest future formation with an inflected form of 'want' in the meaning 'FUTURE' combining with a subordinate verb, e.g., θέλω να λύσω 'I will loosen' (lit., 'I-will that I-loosen'), and the Modern Greek type with the prosodically deficient θα are attested (thus θέλω να λύσω > θέλει να λύσω (impersonal, "it-will that I-loosen") > θέλ να λύσω > θέν να λύσω > θένα λύσω > θάνα λύσω > θάν λύσω > θα λύσω – see Bănescu 1915; Joseph 1983a: chapter 3; Bubenik 2001; Joseph 2001d; Tsangalides 1999; Joseph & Pappas 2002; Joseph 2003b; Markopoulos 2009) and involve Greek-internal sound changes or analogical resolutions of sandhi variants. The same can be said with regard to all the other Indo-European Balkan languages. Therefore, whatever the role of language contact in the emergence of a 'want'-future in the Balkans (and see §6.2.4.1 for fuller discussion), it is hard to see how the prosodic characteristics of the invariant marker could be the direct result of language contact. Presumably, the reductions that led to θα, especially if, as argued in Joseph 2001d, 2003b, θα is best analyzed as an affix and thus prosodically dependent almost by definition, fed into its developing prosodic weakness.[185]

185 It is partly for this reason that we reject the notion of "grammaticalization area" proposed by Heine 1994 and defined in Heine & Kuteva 2005: 182 as "a group of geographically contiguous languages that have undergone the same grammaticalization process as a result of language contact." If language-internal developments underlie the increased grammatical value of some form or collocation, then in what way can language contact be involved or considered responsible for that "grammaticalization"? See Joseph 2011c and §3.4.1.4 for discussion.

In the last analysis, therefore, while there is certainly something prosodically special about the object pronouns, the negation markers, the future markers, the subordinating markers, and so on, invoking language contact as an explanation of the prosodic characteristics remains a difficult approach, except in certain substratum scenarios, and so also with invoking prosody as the basis for convergences, à la Klagstadt. This is not to say that a borrowed item cannot carry some prosodic particularities with it as part of its lexical specifications. The fate of the Macedonian focalizing question marker *li* in Romani is a case in point. In Macedonian (Friedman 2002c), this optional marker appears after the first stressed word in the clause (i.e., after the constituent that occupies focus position) or after the verb, as in (5.30ab), and placement elsewhere is impossible, as in (5.30c); this positioning can be taken as a prosodic characteristic, specifically a prosodic deficiency, since it means that the form does not have the full range of ordering possibilities that a free word has, and this prosodic characteristic is carried along with the form when it is borrowed into the Romani of North Macedonia, as shown in (5.30de):[186]

(5.30) a. Vo Bitola li kje odiš? 'Are you going to *Bitola* (as opposed to Struga)?'
to B. Q FUT go.2SG
b. Vo Bitola kje odiš li? 'Are you *going* to Bitola (as opposed to telephoning)?'
c. *Kje odiš vo Bitola li?
d. šunea li? 'Do you hear?'
hear.2SG Q
e. lake li phučna? 'Are they asking about *her*?'(Bugurdži, Jusuf 1974)
her.DAT Q ask.2SG

With the future marker, by contrast, it is not that foreign material is borrowed, but rather that native material is put to use in an innovative way. For the prosodic similarities to be due to contact, or even for the convergent syntax to be driven by shared prosody, one would have to suppose that special prosodic behavior would have to have been associated with the *concept* of futurity. Even if the deployment of a form of 'want' in a future sense might be at least in part due to contact, one has to wonder how the prosody would be transmitted if no lexical material per se were borrowed. In this way, the future situation contrasts with the borrowing of *li*. For that reason, it seems more likely that the prosodic characteristics were already associated in each language with the reduced future marker, and while the areal distribution in Slavic and Romani, i.e., the geographic restriction (more or less) to the Balkans, would point to contact as a causative factor, it seems that at most contact aided in the selection of that form from amongst a variety of future markings.[187]

186 The same can be said for Turkish *da* 'and' as well as Greek ντε 'expression of impatience' (regardless of whether it is from Slavic or Turkish); see §4.3.4.1.2. Although (5.30c) is not normative, such examples can be found in informal prose on the Internet.

187 See Joseph & Pappas 2001, 2002, and Markopoulos 2009 for some consideration of the range of future markings in Medieval Greek; see also §6.2.4.1.

5.5.4 Intonation

Intonation would seem to be a rich area for study in cases of language contact, as languages do have characteristic intonational contours that might well carry over into the learning of another language. Thus in bilingual individuals and speech communities, intonational overlap between the languages and spreading across the languages involved are to be expected, as indeed is the case in the Romani study (Arvaniti & Adamou 2011) discussed below (see also Queen 1996, 2001, 2012).[188]

Nonetheless, the literature on possible intonational convergence in the Balkans is rather limited. There are several general studies on intonation in individual languages, such as Waring 1976, 1982; Botinis 1998; Arvaniti 2009; Arvaniti & Ladd 2009; Arvaniti et al. 2014; and Baltazani et al. 2019 on Greek, or Penčev 1980 and Miševa 1991 on Bulgarian, and some that treat specific issues, e.g., Dascalu 1975 on yes–no question intonation in Romanian, Grigorova 1998 on question intonation in Romani in Bulgaria, and Grigorova 2001 on Romani intonational formulae. Such studies, however, are essentially on the "suprasegmentals of the Balkans" as opposed to "Balkan suprasegmentals," to draw on and extend the distinction made in §1.2 between languages of the Balkans and Balkan languages, and in §3.2.1.7 between linguistics of the Balkans and Balkan linguistics.

There is, however, some discussion of suprasegmentals in the Balkans in studies that have an eye on contact-related issues and can thus be considered to be examining Balkan suprasegmentals. We discuss these each in turn.

A pilot study of a comparative nature on yes–no questions was carried out by Lehiste & Ivić 1980, examining yes–no question intonation in BCMS and Albanian, augmented with the findings of Dascalu 1975. They state (p. 45) that "there exists no single terminal contour which would characterize all yes-no questions" in BCMS. However, they do report on what they call a "reverse pattern," which they define as follows (*ibid.*): "this tonal contour consists of a low-high sequence and requires for its proper realization at least two successive syllables," noting further that "the location of this tonal contour ... depends on focus or emphasis." In the production of their lone Albanian subject, a man from Kosovo, they found (p. 49) results "quite similar to the Serbocroatian reverse pattern" in that for him "the emphasized word in nonfinal position had a low-high fundamental frequency contour." Since this matched Dascalu's finding of a low–high tonal sequence for Romanian, they venture the very tentative conclusion that this contour is a Balkanism, saying "the answer is a very qualified 'maybe'" (p. 52). Sawicka 1997: 75 is more dubious: "It is hard to tell whether the 'reverse pattern' can be considered a Balkanic feature. It has been established for Balkanic languages, but the same intonation can be heard in Spanish or Italian [and a s]imilar intonation pattern occurs in the yes-or-no questions in Eastern Slavic languages." In the terms

188 We also note that innovative intonations can spread within languages, and thus across dialects, as with so-called "uptalk"(high rising terminal intonation) in various dialects of English (see Guy et al. 1986).

developed here, this putative Balkanism fails to pass the naturalness or universality criterion.

In this regard, too, the findings of perhaps the most comprehensive general work on intonation in the Balkans, Nikolaeva 1996, must be considered tentative, as the author herself recognizes (p. 298). Based on elicitation and readings (of a variety of texts, including poetic ones and folk tales) from speakers of major standard Balkan languages, including also BCMS, this work reports intonation contours that are parallel across the various languages. Moreover, these contours are associated with questions of different sorts, as well as other sentence-types. However, the contour-form and associated function do not really match up across the languages, so that the same contour has different functions in different languages. Also, there is no exact replication of the form and function of Lehiste & Ivić's reverse pattern contour. Nikolaeva does propose one prosodic Balkanism (p. 297), involving high and low positioning on the stressed syllable. However, she recognizes the provisional nature of her suggestion (p. 298), given that it awaits collection and evaluation of similar sentences in other related and unrelated languages. Quite properly, she feels that only by answering such questions can we speak about specifically Balkan phenomena here.

Another suggestive observation is that in Sawicka 1997: 75. She finds cause, especially based on work of her own (Sawicka 1991, 1995), to put forward her own potential prosodic Balkanism. In particular, she describes "a Balkanic tendency . . . [towards] the weakening of the final falling tone and its functional load." She goes on to describe how in colloquial speech in Albanian, Balkan Slavic, and Greek, "statements with a nondistinct final falling tone can often be heard," thereby creating a functional opposition between questions and statements based on utterance-final intonation contours. But her caution suggests that more still needs to be done to make this interesting observation into a more viable claim about Balkan prosodic convergence through contact.

Finally, there is the empirically grounded study of Arvaniti & Adamou 2011 (see also Adamou & Arvaniti 2010). They examine the means for the marking of focus in the Romani of Komotini, in northern Greece (Thrace). Komotini Romani is typologically interesting in having multiple ways of marking focus and using them concurrently. While word order and the particle *da* (borrowed from Turkish) are involved in focus marking, so is prosody, either on its own via accentuation and stress shift or, more usually, in conjunction with other means. What is interesting from a Balkanological point of view is that the Romani community of Komotini is trilingual, in Greek and Turkish as well as their native Romani, and Arvaniti and Adamou argue that Romani's "rich focus marking strategies" may be due to its having "adopted strategies from the languages it has been mostly influenced by, namely Turkish (the use of *da*) and Greek (specific uses of accentuation not common in other Indic languages, such as Hindi and Bengali; Féry 2009)." They conclude that more study is needed, especially of spontaneous speech, but the likelihood of prosodic influence from a co-territorial other language makes this example especially telling. Moreover, it is all the more significant since it seems to

be the result of relatively recent contact, suggesting that the sort of conditions that gave rise to Balkan convergences can still be found today, on a localized basis within the Balkans.

Even with such suggestive results, the study of comparative Balkan intonation is in a nascent state, with much work yet to be done.

5.6 Morphophonemic Alternations

Up to this point in the discussion of Balkan phonology, the emphasis has been on the phonetic realizations of various segments, either synchronically or as the outcome of historical sound changes or as altered in the process of borrowing a word. In addition, some attention has been paid to suprasegmental aspects of pronunciation the various languages.

However, there is more to phonology than segments and prosody and in many instances these realizations, especially when conditioned by specific surrounding sounds, interact with the morphology of the particular languages to produce morphophonemic alternations: changes in the phonemic composition of the form of stems and affixes, whether derivational or inflectional, according to their occurrence in different phonological and morphological environments. The importance of recognizing such alternations was evident in the earliest treatments of Balkan phonology. For instance, one of the features that Miklosich 1862: 7 mentions, i.e., the loss of *l* before *i* in Romanian and Albanian, actually involves a morphophonemic alternation, though it is not labeled as such, in that he contrasts words with final *-l* in one form but with no *-l-* in another related form (e.g., singular vs. plural, as in Alb *portokall* vs. *portokaj* (< **-ał-i*) 'orange' – Miklosich's two examples from Albanian are both no longer current – and Romanian *cel* (< **cel*) vs. *cei* < **celi* 'the, this' (SG vs. PL). Still, the topic of Balkan morphophonemics has not been granted much visibility in the literature.

It can be argued that this lack of attention to morphophonemics is in fact warranted in a study of the commonalities amongst the Balkan languages because all of the Balkan languages present a wide variety of morphophonemic alternations, mostly involving inherited material with so little actual overlap across the languages that there is little of significance to report on. Most of these are really just phenomena restricted to individual languages ("*einzelsprachliche*") and so are not comparable across the languages. For instance, within the history of Postclassical Greek, a sound change occurred by which κ became χ when it occurred before τ, as in οχτώ 'eight' from earlier ὀκτώ. While this had the effect of changing the lexical form of some words, as with 'eight,' in other instances it meant that there was a new alternation in the form of particular morphemes, as with the root πλεκ- 'knit' which has a *κ* before a vowel-initial suffix, e.g., πλέκ-ω 'I knit,' but a χ before a suffix beginning with -τ-, e.g., πλεχ-τό 'wicker.' This sound change and the resulting morphophonemic alternation it occasioned are interesting for the history of Greek but they have no parallel within any Balkan language; such examples of language-particular

morphophonemics can be multiplied across the Balkans. Providing a list of developments of this sort would only demonstrate that Balkan languages are like numerous other languages of the world in having morphophonemic alternations. Thus, any such alternations mentioned would be of little value or interest in terms of Balkan linguistics as defined in §3.3, i.e., as a contact phenomenon within the sprachbund, though they would be relevant for a linguistics of the Balkans, i.e., an accounting of the facts of particular languages viewed as independent entities. As is discussed below, some morphophonemics are affected by or even introduced when speakers of different languages are in contact with one another, but much of what the Balkan languages have in the way of morphophonemic alternations does not involve language contact, making these features necessarily not of interest from the Balkanological standpoint.

It is also sometimes the case that the triggering conditions for morphophonemic alternations involve phonologically natural processes, as with assimilations between adjacent segments. In such instances, unless there is a compelling reason, perhaps involving an unexpected or seemingly unnatural environment for the alternation that is duplicated in each language, it generally can be argued that the alternations are likely to have arisen independently in each language.

Still, there are some parallels in Balkan morphophonemics that can be noted, and some seem, or at least have been claimed, to involve language contact. Accordingly, we discuss a few of them here, even though ultimately our conclusion about most instances of Balkan parallels in morphophonemics is that they do not require one to look to language contact for an explanation, except, perhaps, when they are leveled out or when borrowed material is involved.

One case involves palatalization in Romanian and the possibility of Slavic influence being responsible for its emergence, as proposed especially by Petrovici 1956, 1957, 1958. Although this is as much an issue for the segmental inventory and set of phonological oppositions in Romanian as it is for morphophonemics,[189] we address it here from the standpoint of morphophonemic alternations. Petrovici claims that the palatalization found in Romanian masculine plurals, e.g., *cipic* ([čipik]) 'boot lace' / *cipici* ([čipič]) 'boot laces,' or *băiet* ([băjet]) 'boy' / *băieţi* ([băjet^{s}]) 'boys,' is to be compared with similar alternations in Bulgarian, e.g., *ezik* 'language' / *ezit^{s}i* 'languages.' The Bulgarian alternations are the reflexes of the so-called Second Palatalization of Common Slavic (Scatton 1993: 194) and are thus known to be old, inviting a suggestion of Slavic influence on Romanian in the area of morphophonemics. However, as Petrucci 1999: 48–49 points out, the range of consonants affected and the specific outcomes of the palatalization differ between the two languages, diminishing the likelihood of

189 Petrucci 1999: 41–49 gives a useful summary of Petrovici's analysis, which posits distinctive palatalized, labialized, and palato-labialized segments in Romanian alongside neutral ones. Petrucci also offers a convincing critique of Petrovici's system, rejecting it as typologically and phonetically ill-supported and as adding complications to otherwise straightforward processes of Romanian. Without such an analysis and consonantal inventory, a claim of such systemic Slavic influence on Romanian, Petrucci concludes, is hard to support (though he approves of several individual cases of Slavic influence).

a Slavic origin for the Romanian developments. Moreover, any change involving palatalizations triggered by front vowels must be judged as phonetically very natural and certainly cross-linguistically quite common. Thus looking to language contact as the source of these Romanian morphophonemics is not particularly compelling; they could easily have developed in Romanian on their own, as natural phonological processes. We can add that these processes are found in fact in Balkan Slavic and Balkan Romance more generally, not just in Bulgarian and Romanian.

Second, as discussed above in §5.4.5.2, word-final devoicing, for the languages that have it, is a source of some different morphemic shapes.[190] In the typical case, a morpheme that undergoes final devoicing occurs in some other contexts in which devoicing does not occur. For instance, in Macedonian the indefinite form for 'bread' is pronounced [lep], but in the definite form it is realized with a stem-final *-b-*, *leb-ot* 'the bread.' Similar alternations are found in Bulgarian, northern Tosk Albanian, Turkish, and so on. To the extent that such final devoicing is a contact-induced feature that spread from an original locus in Macedonian into contiguous areas in Albanian and South Montenegrin and affected some Romani and even one Greek dialect in the central Balkan region, one can say that the morphophonemic alternations that have resulted in these languages are also the product of language contact. However, it should be clear that it is not the morphophonemics themselves that are borrowed but rather what is borrowed is the contextually restricted pronunciation adjustment of devoicing an obstruent in word-final position. Once such a sound alteration enters a language, then, assuming the morpheme in question can occur with the relevant obstruent in other than word-final position, there automatically arises a voiceless/voiced alternation of the sort that would be captured formally by some mechanism relating the two pronunciations of the morpheme in question. Language contact is thus involved in the parallel morphophonemic alternations in these languages, but only indirectly, by providing a particular pronunciation to a morpheme in only one context out of the many in which that morpheme can occur.

Third, as observed in §5.5, a striking feature of Macedonian is the relative simplification of morphophonemic alternations vis-à-vis Bulgarian. An obvious fact is that western Macedonian has fixed stress, and that as one moves eastward toward Bulgarian, i.e., away from the most intense Balkan multilingual complexity represented by western Macedonia, Slavic morphophonemic accentual alternations increase in complexity. But another fact worth mentioning is that Macedonian has considerably fewer morphophonemic – and even morphological – alternations than does Bulgarian in their respective verb paradigms, and again southwestern Macedonian represents the most extreme outcome of simplification. As argued by Markovikj 2017, these regularizations are connected with the mutual

190 In this case, the morphemic shapes are the result of automatic phonology in the languages in question; in other instances, the varying shapes are governed by nonphonological factors (e.g., grammatical category).

multilingualism of the region, and see below regarding Trudgill's views on such matters and also Elson 1995.

Finally, there is one morphophonemic feature found in parallel in three languages or groups in the Balkans, with there possibly being a fourth as well, that has not been mentioned in the literature on the Balkans (though see Joseph 2007b). In one sense, this lack is perhaps surprising since the feature is a well-known and somewhat striking one, but in another sense also it is not surprising because its history in each language involved is so well known that the feature is an obvious language-independent development in each language. Still, we mention it here so as to cover all potentially relevant territory, and also for the methodological point it makes concerning the distinction between areal causation vs. typological resemblance. This is the case of so-called "fleeting vowels" in Balkan (and the rest of) Slavic, in Turkish, and in Albanian, with northern Greek dialects forming the possible fourth entry.

In these languages, there are words in which vowels occur in final closed syllables that are absent in related forms with a different syllable structure. Thus in Turkish the stem for 'city' is *şehir* when the *r* closes a syllable, e.g., in the nominative singular *şehir* or the nominative plural *şehir-ler*, but *şehr* when the *r* is a syllable onset, as in the dative singular *şehr-e* 'to the city' or the definite accusative *şehr-i* 'the city'; similarly, an example from Bulgarian is the stem for 'silver', *srebăr-* when the *r* closes a syllable, as in the derived noun *srebăr-nik* 'silver coin,' but *srebr-* when the *r* is an onset, as in the base noun *srebr-o*.[191] And, in Albanian, the noun *emër* 'name' occurs in that form in the nominative singular indefinite form, but with the definite marker, the form is *emr-i* 'the name.'[192] What Greek adds here is forms in northern dialects, where unstressed high vowels are subject to loss (see §5.4.1.5) and alternations in stress placement create "fleeting vowels" in different forms of the same word, e.g., μάθμα (vs. standard (and southern) μάθημα) 'lesson.SG.NOM.ACC' / μαθήματα 'lesson.PL.NOM.ACC'; since these Greek alternations are tied to the presence/absence of stress on the vowel in question, they may be counted as different altogether from the Albanian, Balkan Slavic, and Turkish phenomenon.

Thus the same V~Ø morphophonemic alternation, occurring under similar, largely syllabically based, conditions, is found in three branches within the Balkans. Since in some instances, less of a superficial convergence than that has been reckoned as a possible Balkanism, it is interesting, and instructive, to consider this one. This parallel, however, as suggested above, is nothing more than a mirage as far as Balkan language contact is concerned, and thus makes an important point

191 The statement of the distribution of V/Ø in Bulgarian is actually more complicated, as there are formations in which the potential for vowel loss/insertion (depending on how the alternation is formalized) occurs in two successive syllables; thus the adjective 'silver,' formed with the suffix *-n-*, has the masculine singular form *srebăren*, but feminine singular *srebărna*; the selection of the allomorph *srebăr-*, therefore, depends on the vowelless form of the adjectival suffix. The same applies, *mutatis mutandis*, for the rest of Slavic.

192 This statement oversimplifies matters as there are similar words where, at least as far as their canonical form is concerned, *ë* does not disappear, e.g., the related verb *emëroj* 'I name.'

about methodology that history is essential to measuring the value of a typological parallelism (see Joseph 2007b for some discussion).

In the case of Balkan Slavic, it results from several sound changes: the loss of the Common Slavic *jers* (ultra-short high front and back vowels) in so-called weak positions as opposed to their preservation (ultimately giving *e/e/ă* and *ă/o/ă* in Standard Bulgarian/Standard Macedonian/Torlak BCMS, respectively) in strong positions.[193] These vowels (and sometimes others) also show up as so-called secondary jers to break up consonant clusters resulting from the loss of jers, e.g., OCS *reklъ/rekla* 'said (resultative PST.PTCP.M.F),' Mac *rekol/rekla*, Blg *rekăl/rekla*, BCMS *rekao/rekla*. Cf. Scatton's 1993: 193 formulation of "epenthetic *ă*," the historical insertion of *ă* "to break up stem-final consonant clusters terminating in liquids or nasals." And, similar alternations are found throughout Slavic more generally, so the impetus behind the developments leading to the Bulgarian alternations (and similar ones elsewhere in Balkan Slavic and beyond) is to be located in the break-up of Common Slavic. As for Turkish, however, the words with the "fleeting vowel" are mostly borrowings from Arabic with word-final clusters that were otherwise not found in Turkish that were nativized in the borrowing process by the insertion of an epenthetic vowel. And, as far as Albanian is concerned, the alternation seems simply to be the result of a rather natural change, loss of an unstressed vowel word-medially, that is not only quite common cross-linguistically (note English disyllabic pronunciations like [fæmli] for canonically trisyllabic *family*) but also quite widespread as a fast-speech reduction in the language (note, for instance, futures with the *ë* of the modal subordinator elided, e.g., *do t'shkoj* 'I will go'). Thus the Slavic developments, the Albanian developments, and the Turkish developments have nothing to do with one another historically. As in all instances of working out the importance of structural parallels in the Balkans, history prevails, no matter how enticing it might be to look for a connection between two seemingly similar phenomena.

There is however, one instance in Albanian where contact with Slavic does seem to play a role in this feature, namely in the transformation of schwa to /o/ as the Albanian weak vowel in Opoja Geg Albanian (Dombrowski 2012, 2013: 106–114). Opoja (Slv Opole) is the region just north of Gora and south of Prizren, and the region shifted from Slavic to Albanian at some point. Opoja Geg is unique in Albanian in having /o/ as the weak vowel. Dombrowski writes:

> The location of Opoja on the boundary separating two types of Slavic jer reflexes (/ə/ [for both jers] to the north and /o/, /e/ [for strong back and front jer respectively] to the south) facilitated the identification of Albanian /ə/ with the Slavic jer reflexes. [. . .] If . . . Opoja [Slavic] had only /o/ from the jers, its imposition onto **ə* follows automatically. Such an account would involve the

193 This much-discussed distinction, usually associated with the label *Havlik's Law*, can be stated as follows (adapted from Schenker 1993: 78): "*jers* were weak in word-final position, strong before [a syllable containing] a weak *jer*, and weak before [a syllable containing] a strong *jer* or any other vowel." There are many exceptions to this generalization, but the basics are consistent enough to serve the purposes here.

direct imposition of Slavic (morpho)phonology onto Albanian during the period of language shift, and is much simpler than a process in which one of two alternating vowels was chosen.

As the Turkish case shows, too, borrowed lexical material can be involved in the emergence of morphophonemic alternations in a language. Arabic loanwords in Turkish presented a phonological problem in some forms but not others, and resolving that problem created morphophonemic alternations. The involvement of language contact in the Turkish fleeting vowel alternations, therefore, was indirect at best; the development did not take place until after the relevant forms entered Turkish.

There are other ways that borrowed material figures in morphophonemics, though upon closer inspection, not all turn out to be a proper concern here. We start with the assumption (see §3.2.1.7 and §4.2.2) that in cases where foreign affixes enter a language, they generally are not borrowed per se; rather they enter in whole words, attached to particular borrowed stems. If the borrowing of such whole words brings two different forms of the same source-language affix into the borrowing language, then the borrowing language necessarily retains the morphophonemics of the source language for that form, whether or not the alternations conform to regular phonological patterns in that language. Thus when Albanian and Balkan Slavic borrowed Turkish words complete with Turkish plural morphology, as in Albanian *baba-llarë* 'fathers' (Turkish *babalar*) or *bej-lerë* 'Turkish notables' (Turkish *beyler*), Macedonian *kardaš-lar* 'brothers' (Turkish *kardaş-lar*) or *beg-ler-i* (Turkish *beyler*, with a Macedonian plural ending added – the final *ë* in Albanian is likewise a native plural marker), the languages came to have an alternation in the shape of a suffix that mirrored the Turkish distribution of vowels, with suffixal *-a-* after back-vowel roots and suffixal *-e-* after front-vowel roots.[194] This harmonic alternation is alien to Albanian and Balkan Slavic, being restricted (mostly) to just these few lexical items of Turkish origin, but it is, nonetheless, an alternation which a full accounting of the morphophonology of these languages would need to describe.[195] Regardless of how it is formalized, any account of the Albanian and Balkan Slavic *-a-* / *-e-* alternation in this highly restricted suffix would resemble, on a much more limited scale, the formalization of the alternation in Turkish. Rather than claiming that the Turkish rule of vowel harmony was borrowed into these languages, it seems best to say that the material from which one would infer or induce such a rule was borrowed, and that the particular rule formulation occurred within the individual borrowing languages; the parallelism between this very limited vowel harmony in Albanian and Balkan

194 The quality of velar vs. palatal /l/ is automatic before back and front vowels, respectively, in Balkan Slavic, and after them in Standard Turkish, but the Turkish automatic alternation is realized as distinct phonemes in Albanian. See §6.1.4.1 for details on Turkish plurals in various Balkan languages, including a motivation for the borrowing in terms of the prestige associated with the language and the nouns involved.

195 There are some extensions attested in Albanian of the Turkish plural to non-Turkish stems; see §6.1.4.1.

Slavic and the more widespread and productive process in Turkish is at best indirect.[196]

Also, a borrowed element can develop morphophonemic alternations after it has been borrowed due to sound changes or other processes in the borrowing language; if those alternations mirror ones found in the source language, then by accident one can have the appearance of the spread of the alternations via language contact. And in some cases it simply is too hard to tell. An example is what has happened with the Turkish agentive suffix *-CI* in several of the Balkan languages (cf. §4.2.2.4). This suffix in Turkish has two forms as to its initial consonant: voiced *-CI* after voiced segments and voiceless *-çI* after voiceless.[197] In some of the languages that have borrowed *-CI*, both voiced and voiceless realizations occur, depending on the stem-final environment. In Greek, for instance, one finds both μπογια-τζής 'painter,' derived from μπογιά 'paint,' and καϊκ-τσής 'owner of a small boat,' derived from καΐκι 'small boat'; since the base nouns are both borrowings from Turkish (cf. *boya* 'paint' and *kayık* 'boat'), it is likely that the τζ/τσ alternation reflects the pronunciation of the associated agentive nouns in Turkish itself. However, this alternation is found as well with some neologistic uses, where the words could not have been borrowed from Turkish, so that they are clear cases of the innovative use of the suffix within Greek itself with the relevant alternation. Thus, one finds both ταξι-τζής 'taxi-driver,' with voiced -τζ- in a voiced context (after the final vowel of the base noun ταξι-) and ΠΑΣΟΚ-τσής 'an adherent of the ΠΑΣΟΚ political party,' with voiceless -τσ- in a voiceless context (after the final *-k-* of ΠΑΣΟΚ). What is difficult about assessing this alternation and the role of language contact here, as in the other Balkan languages with *-CI*,[198] is that Greek (as well as the other languages, and, for that matter, numerous other languages of the world more generally) does not tolerate voicing mismatches in clusters involving obstruents.[199] As a result, certainly for the neologistic uses, but possibly also for the older Turkish-derived vocabulary, it seems best to treat the Greek -τζ-/-τσ- alternation in this suffix as caused by Greek-internal phonological processes. In such a case, then, even with replication in the borrowing language of source language morphophonemics, there is no way – given that natural phonological processes present anyway in the borrowing language are at work – that one can state positively that direct borrowing of a foreign language's morphophonemics has

196 See §5.4.3.7 above concerning other mentions of vowel harmony in the literature on Balkan languages.

197 We use <C> as a cover symbol for the consonant that surfaces as either voiceless *ç* or voiced *c*, and <I> for the high harmonic vowel, which is irrelevant for Greek but surfaces in Turkish as [i ɪ ü u] depending on the root vocalism, although in WRT <I> is realized only as /i/ in final position. As noted below, the statement of the distribution of <c/ç> with this suffix is more complicated in Turkish than stated here since devoicing occurs in somewhat unexpected contexts.

198 For example, Macedonian has *sladolet-čija* ice-cream dealer' but also *lov-dzija*, 'hunter' and Albanian has *batak-çi* 'deadbeat' vs. *kompromisa-xhi* 'boot-licker.'

199 An apparent exception to this in the Balkans is Albanian *bixhozçi* 'gambler' (Kostallari 1976: s.v.) from *bixhoz* 'a gambling game, usually with cards' derived ultimately from *bixhas* 'a kind of tic-tac-toe played with stones' (Çabej 1976: s.v., who, however, gives *bixhozxhi*).

occurred (see Karaśinski 2018 on the details of voicing issues with this suffix in the various Balkan languages).

If however some less-than-natural detail about the morphophonemics is replicated in a borrowing language, then one is in a better position to treat the morphophonemic parallel as being borrowed in the abstract; looking to the more concrete realization of particular morphemes in particular words in the borrowing language would seem to be more difficult to justify. Such may be the case with one aspect of the borrowing of *-CI*. In particular, in Turkish, the voicing of *-CI* depends not on underlying voicing of the root-final segment but on the form the root-final consonant takes in combination with this suffix, where a process of syllable-coda obstruent devoicing can take effect; thus *kitab-* 'book' ends underlyingly in a voiced stop, but with the agentive suffix, one finds *kitapçı* 'bookseller,' with devoicing of the root-final labial stop and thus selection of the agentive allomorph with *-ç-*. This "post-lexical" attachment of *-CI* onto what amounts to the surface form of the base noun – note that *kitap* is the form this stem takes when occurring freely – as opposed to its underlying form, though perfectly regular as far as the attachment of Turkish suffixes is concerned,[200] can be considered somewhat unusual cross-linguistically for an element that appears to be affix-like in other respects. Interestingly in Macedonian, the same sort of post-lexical attachment of *-džija* is found, with devoicing of the root-final element and suffixal consonant just as in Turkish; thus the stem *sladoled-* 'ice cream' (with an indefinite form [sladolet] but a definite form [sladoled-ot]) forms an agentive noun *sladoletčija* 'ice-cream seller.' In such a case, one might feel more justified in thinking in terms of the borrowing of the morphophonemics per se, i.e., of some more abstract property of the agentive marker, rather than the extraction of that property out of borrowed full-word forms containing the marker. However, even there one could start with the behavior of the borrowed element in some particular words in which it was borrowed, and work outward from there, especially if this devoicing property originally occurred in the usage of Macedonian speakers very familiar with Turkish and spread from that locus of innovation to monolingual speakers of Macedonian.

There is one final point to be made about morphophonemic alternations in the various Balkan languages and language contact. That is, besides the entry of (the potential for) morphophonemic alternations through borrowed elements, contact can also be involved in the loss of alternations in a given language. Alternations can be lost in purely language-internal ways via (analogical) leveling, in particular if one allomorph is generalized at the expense of another. Thus in Greek, the alternation between [b] after nasals and [p] elsewhere, as in την πόρτα ([tim borta]) 'the.ACC door.ACC' but η πόρτα ([i porta]) 'the.NOM door.NOM,' has been leveled in some dialects in favor of [b], giving η μπόρτα ([i borta]) (e.g., in Cretan).[201] If the

200 For instance, the ablative of 'book' is *kitap-tan*, and the locative is *kitap-ta*.

201 Leveling in the opposite direction is found too; Arvaniti & Joseph 2000 report that some speakers of Athenian Greek have [ti pira] (τη πήρα) 'I took her' for expected [tim bira] (την πήρα), presumably based on the form with initial [p-] that occurs in nonnasal environments (e.g., [to pira] (το πήρα) 'I took it'). This might also be thought of as hypercorrection (cf. Kazazis 1968 on such so-called Sunday Greek).

leveling occurs in a sufficient number of lexical items that could, or once did, show the alternation, then the alternation would disappear from the language altogether.

However, as shown in §3.2.1.5, language contact, according to Trudgill 2006 and as indicated above, can play a role in the loss of an alternation via leveling in a contact situation in which there is widespread second-language learning by adults. Under this view, an interesting difference between Macedonian and Bulgarian may find an explanation. In particular, it has been observed (e.g., by Elson 1995) that many morphophonemic alternations that are found in Bulgarian have been levelled out in Macedonian; for instance, corresponding to Bulgarian *piše-* (present) / *pisa-* (aorist), Macedonian has *piša-*/*piša-*, with the consonant alternation eliminated. In Trudgill's model, this fact would be explainable by the greater degree of language contact in North Macedonia than in Bulgaria; that is, there is a more linguistically diverse population in North Macedonia than in Bulgaria so that over the years, there has been a higher percentage of second-language learning of Macedonian in the overall speech community than of Bulgarian. Markovikj 2017, cited above, makes a similar point.

Overall, then, while there is perhaps less in the way of contact-induced developments with morphophonemics than with other aspects of phonology, it is not an area totally devoid of language contact effects.

5.7 Expressive Phonology

For the most part, linguists tend to view sounds as just the incidental material that morphemes and words are made of. The sounds of a language in and of themselves are important, to be sure, and form a coherent subsystem within a language overall, complete with its own properties of internal relations and such. However, their existence is manifest only in their occurrence in particular morphemes and words. Yet there are situations in which sounds alone can be carriers of meaning. Typically, in such cases, the sort of meaning involved is more affective and attitudinal, and is not essential to the function of language as a means for communication among fully competent users. This is not to say that language is only for adults talking to adults for purely informational purposes, but rather to suggest that the practice among many linguists is to focus on those aspects of language. Here, though, we draw attention to the classes of lexical items, typically onomatopoeia, sound symbolic words, ideophonic expressions, and the like, in which such meaning-bearing sounds, which we refer to as *expressive phonology*, can be found. These lexical categories are highly expressive, and to a large extent make language fun, give it life and color, and allow for individuality in expressiveness.[202]

202 One might refer to such words as "language with an attitude" (a phrase used differently by Preston 2002, as a title to lure readers to a discussion of folk attitudes about language and about speakers of particular languages and dialects).

Expressive phonological phenomena are to be found in the languages of the Balkans, and, not surprisingly, there are some convergences to be observed in this domain. Expressive phonology in Balkan languages has been studied most extensively for Greek, primarily by Joseph in a series of articles (Joseph 1982, 1983b, 1984ab, 1987c, 1994b, 1997b; see also Joseph & Philippaki-Warburton 1987: 258–261) in which there is some reference as well to what is found in languages neighboring on Greek. As this is a relatively unexplored area of Balkan phonology, the discussion here takes Greek as the point of departure and focuses mainly on findings for that language, with some reference to other languages where feasible. Expressive phenomena of course are not restricted to phonology; some aspects of Balkan expressivity are more morphological or morpholexical in nature, and as such are treated in Chapter 4 – see, e.g., §4.3.6 on onomatopoeia, §4.3.7 on reduplication, and §4.3.8 on diminutives, hypocoristics, and endearing terms of address.

The basic observation about phonological expressivity in Greek is that it seems to be centered on two sounds, the voiceless and voiced dental affricates τσ and τζ.[203] This is not to say that no other sounds or phonological processes figure in expressivity – for instance, as noted in §5.4.1.1, ii, vowel lengthening in Greek can have a discourse function as a hesitation marker; however, a key fact is that there is an extensive network of forms and uses involving τσ and τζ.

The basic relevant fact about Greek τσ and τζ is their lexical distribution; their occurrences are skewed in that they are found for the most part in forms that fall into lexical categories that can only be characterized as marginal. For instance, the lexical classes in question include those containing sound symbolic combinations (5.31a), various widespread diminutive formations (5.31b), interjections (5.31c), calls to animals (5.31d), onomatopoeia and derivative formations (5.31e), ideophonic adverbial expressions (5.31f), conventionalized forms used by adults to and around children (5.31g), and a wide range of words that are colorful, playful, expressive, and in general somewhat slangy, ones that in short lend color to language in ways that go beyond simple conveyance of some denotational sense (5.31h); given the extensiveness of voiceless τσ in these classes, that sound is the focus here, but relevant cases with voiced τζ are mentioned where appropriate (relevant elements in these forms are highlighted in bold:

(5.31) a. sound symbols, e.g.: **τσι-** ‘small, narrow, thin,’ as in **τσι**τώνω ‘stretch,’ **τσί**χλα ‘thin woman,’ **τσί**ρος ‘thin person,’ **τσί**τα-**τσί**τα ‘just, barely (said of a narrow squeeze or a tight fit)’;[204] **τσV-** ‘sting, tease, bite, burn,’ as in **τσού**ζω ‘sting,’ **τσου**κνίδα ‘nettle,’ **τσι**μπούρι/**τσι**μμούρι ‘tick’ (“small stinging insect”), **τσι**βίκι ‘tick,’ **τσι**μπώ ‘pinch,’ **τσα**τίζω ‘I tease,’ **τσιτσι**ρίζω ‘sizzle, torment slowly,’ **τσου**ρουφλίζω ‘singe,’ **τσί**κνα ‘smell of meat or hair burning’

203 See footnote 55 and §5.4.6 regarding the treatment of Greek τσ and τζ as affricates.

204 The Turkish borrowing **τζου**τζές ‘dwarf’ may belong here too, thus extending the “family” of sound symbols to include the voiced counterpart to τσ; see also footnotes 206, 207, 208. On the other hand, given the lack of palatal affricates in much of Greek and their substitution with dental affricates, the full extent of expressivity of τζ is an open question.

b. diminutives with nucleus **-τσ-** (cf. also **τσι-** 'small' sound symbol): neuter diminutive noun suffix **-ιτσι**, as in κορί**τσι** (cf. κόρη 'girl, daughter'); feminine diminutive noun suffix **-ίτσα**,[205] as in λεμονί**τσα** 'little lemon tree' (cf. λεμονιά 'lemon tree'); nonsuffixed hypocoristics derived directly from names, as in Μή**τσος** (from Δημήτριος) and Κώ**τσος** (from Κωνσταντίνος); "diminished" adjectives, as in γλυκού**τσικος** 'sweet-ish, cute' (cf. γλυκός 'sweet') or καλού**τσικος** 'good-ish' (cf. καλός 'good')[206]
c. interjections, e.g., πρι**τς** 'so what?!; who cares?!,' **τσα** 'noise used in peek-a-boo game.'[207] Orthographic <τσ> 'no' represents an ingressive voiceless dental affricate (alveolar click), i.e., a different sound that is not otherwise part of the phonemic system (see §4.3.3.3).
d. calls to animals, e.g., γού**τς** 'call to pigs,' **τσ**ού(νκ)ς 'call to donkeys,' ό**τς** 'whoa!,' ί**τς** 'whoa!'
e. onomatopes and derivatives, e.g., **τσ**ακ 'crack!' (cf. **τσ**ακίζω 'I break'), κρι**τς**-κρι**τς** 'crunch!' (cf. κρι**τσ**ανίζω 'I crunch'), μα**τς** μου**τς** 'kissing noise,' **τσ**ιου-**τσ**ιου 'bird's chirp,' πλι**τς**-πλα**τς** 'splish-splash!,' γρα**τς** 'scratching sound' (with variants χρα**τς**, κρα**τς**, and κρα**τς** κρου**τς**, and derivative γρα**τσ**ουνίζω 'I scratch')
f. ideophonic adverbials (where the sound is evocative of a manner of action), e.g., **τσ**άκα **τσ**άκα' immediate quick action; straightaway; directly,' **τσ**ούκου **τσ**ούκου 'steadily and surely, with a hint of secretive activity,' **τσ**άφ **τσ**ούφ 'in an instant'
g. adult conventionalized child-language forms, e.g., **τσ**άτσα 'aunty,' **τσιτσ**ί 'meat' (also adult slang for 'breast'), **τσ**ίσ(ι)α 'peepee,' πί**τσι** πί**τσι** '(act of) washing'
h. expressive, playful, slangy words, e.g.: **τσ**αμπουνίζω 'whimper; prate; bullshit,' **τσ**αλαβουτώ 'do a slovenly job,' **τσ**όκαρο 'vulgar woman' (primary meaning: 'wooden shoe'), **τσ**ιρίζω 'screech', **τσ**ιλημπουρδώ 'gallivant; fart about', **τσιτσ**ίδι '(stark) naked.'[208]

Greek τσ can also occur in various perfectly ordinary words with no marginality to them at all, some native, such as έτσι 'so, thus,' and some loanwords (with dental τσ for palatals in the source language, such as τσιμέντο 'cement' (cf. Itn *cimento*), or παπούτσι 'shoe' (cf. Trk *papuç*) but the preponderance of its lexical occurrences is overwhelmingly in words such as those in (5.31) other than in loans from languages with palatal affricates.[209]

205 This suffix is especially widespread in the Balkans, and also the subject of some controversy as to its origin, as discussed in §4.3.8; see also Joseph 2015b.

206 The suffix **-τζί**κο-, which forms affective adjectives (or perhaps, adjectives from (certain) affective nouns), as in μασκαρα**τζί**κος 'scoundrel-like, clearly dastardly' (cf. μασκαράς 'scoundrel'), may belong here, extending the formal side of this group much as adding **τζ**ουτζές 'dwarf' extends the τσ-based sound symbols (see footnote 204); see also footnotes 207, 208, 209.

207 A variant, and perhaps more common, form of this last element is **τζ**α (see also footnotes 204, 206, and 208, as well as 215 on an Albanian parallel); see also §4.3.4.3.2 and §4.3.10.1.2.2 on its use in the game of peek-a-boo. Note too that Macedonian shares the use of *dza* as the "I see you!" expression.

208 Here too τζ words can be added (see also footnotes 204, 206, 207, 209): **τζ**άμπα 'for free; cheap' (from Trk *caba*), **τζ**ιριτζάντζουλες 'evasiveness, coquettish airs,' and **τζ**άντζαλα μάντζαλα 'rags and such' (this last with the expressive *m*-reduplication from Turkish – see §4.3.7). All of these, however, depend on the fact that (Standard) Greek lacks palatal affricates.

209 So also for τζ, as in τζάμι 'glass window' or τζαμί 'mosque.' τσ and τζ are the two least frequent sounds in Greek; see Mirambel 1959; Householder et al.1964; and Joseph 1994a.

Interestingly, the marginality of τσ and τζ, as seen in part in their skewed lexical distribution, is mirrored iconically in some further classes of words that contain these sounds. That is, these marginal sounds occur in various words referring to marginal social groups, as defined by the characterization of physical deformities and thus referring to people who are physically marginalized, as in (5.32a); in descriptions of negative character traits, which often place an individual at the margins of society, as in (5.32b); and, coincidentally, in words for Roms and Jews, the two most marginalized groups both living in isolation from the mainstream of Greek society, as in (5.32c). However, the words themselves are usually loans from languages with palatal affricates, thus the marginalization is doubled, since Greek linguistic ideology abhors loanwords:

(5.32) a. (-)τσ- in words for various deformities: **τσ**ευδός 'lisping; stammering,' **τσ**άτρα πάτρα 'stumblingly (especially of speech)' (see §4.3.7.3), κου**τσ**ός 'lame,' κα**τσί**δα 'balding, scurvied head,' κα**τσ**ο- 'wrinkledy-,' **τσ**ιμπλιάρης 'bleary-eyed'

b. #τσ in words for character flaws or negative traits: **τσ**απατσούλης 'slovenly,' τσουλής 'untidy person,' **τσ**ούλα 'loose-living or low-class woman; slut'

c. (-)τσ- in certain ethnic labels: **τσ**ιγγάνος (also αθίγγανος, κα**τσ**ίβελος) 'Rom', cf. **τσ**ιγγούνης 'miserly;[210] **τσ**ιφούτης 'skinflint' (originally from Trk *Çıfıt* 'Jew')[211]

While expressive phonology, as noted above, in principle need not be restricted to affricates,[212] it happens that these sounds figure in other Balkan languages in ways reminiscent of the Greek case, though perhaps not so thoroughly developed or elaborated as in Greek, or not as restricted to such usages as in Greek, since they all have palatal affricates available.

Marchand 1953: 59, for instance, has noted that the Turkish voiceless and voiced palatal affricates *ç*/*c* serve a similar function to the Greek τσ/τζ, occurring, as summarized by Joseph 1984b: 233, "in numerous words which he terms "*lautsymbolisch"* [('sound symbolic')], including words for murmured and vibrating noises, words of 'affective' origin, and the like … [Moreover,] he notes onomatopoetic

210 Although Byzantine Greek ἀθίγγανος 'untouchable' is often cited as the etymon of Grk τσιγγάν-, there are serious problems both phonological and semantic; Matras 2011 argues convincingly for OT *čiγan* 'low caste person' being the source. See also §4.3.9.4 and Chapter 4, footnotes 109, 310.

211 Turkish *Çıfıt/Çıfıt* is nativized from learned *Cuhud/Cühud* 'Jew' from Pers *Cuhūd* 'idem,' ultimately from Heb *Yāhūdī* 'member of the tribe of Judah > Jew,' possibly via an Arabic intermediary (Tietze 2002: s.v.). The association of undesirable traits with marginalized ethnic groups is, alas, universal, although the assignments of specific traits varies, e.g., sneaky vs. stupid, violent vs. cowardly, stingy vs. spendthrift, loose vs. uptight, as well as unpaired negative traits like lazy and dirty, etc.; cf. Roback 1944. The Greek form τσιφούτης is now so opaque to modern Greeks that they are unaware of its origin as an ethnic slur, much as most Americans do not know that *gyp* 'cheat' is from *Gypsy* or much as Christians in the southeastern United States use *jew* as a verb meaning 'bargain, haggle' without necessarily making the connection to a religion they regard as alien.

212 Cross-linguistically, coronal affricates like [ts] and [č] often occur in diminutive formations or are associated with sound symbolic sequences denoting smallness, so that this recurring association may well be due to their higher pitch (high second formant), as high pitch is characteristic of the noise made by small objects. See Hinton et al. 1994 for some discussion, and also Joseph 1984a for some comparisons outside of the Balkans.

forms like *cıvıl-* 'twitter,' affective pet names like *cici*, [and] conventionalized child-language forms like *çiş* 'peepee.'" To this can be added the fact that some of the relatively few lexical occurrences of the quite rare *dž* in Bulgarian are onomatopoeia, and the *dž* in Macedonian, as a variant of *ž*, can have, as Friedman 2002c: 10 puts it, an "expressive effect." With regard to initial *dz* in Balkan Slavic, it can be noted that in Bojanova 1998, a Bulgarian–English dictionary with 90,000 headwords, only two begin with *dz*: *dzăn* 'ting(a-ling)' and *dzift* 'pitch, bitumen, asphalt.' BER I has 114 words with initial *dz,* most of them Macedonian (in keeping with Bulgarian politics), the remainder being onomatopoetic or expressive and dialectal. In Macedonian, the affrication of *z* to *dz* is a western feature that became part of the standard, e.g., *dzid* 'wall' (from older *zid*). This phenomenon was clearly not necessarily expressive. However, in initial position it is limited to about a half-dozen lexical roots plus some onomatopoeia, and of the 1023 pages of Murgoski's 2013 encyclopedic Macedonian–English dictionary, totaling 300,000 entries, initial *dz* takes up just one and a half pages.

A similar situation obtains in Albanian, where, according to Curtis 2010, the strident palatal voiced affricate spelled <xh> is the least frequent phoneme in a phoneme frequency count, coming in at only .02 percent occurrence, and its dental counterpart, spelled <x>, is the next least frequent at .03 percent. In fact, Curtis argues for an expressive status in Albanian, similar to that developed here for Greek, for <xh>, based on its common occurrence in onomatopoeia and expressive, sometimes reduplicative, formations, as in (5.33), taken from Newmark 1998:[213]

(5.33) *xhagajdur* 'cocky braggard who goes around looking for a fight: bully'
xhahil '(person) who is ignorant, backward, uncultured and thickheaded'
xhambaz 'swindler, con-artist'
xhaxhi '[children's usage] term of affectionate respect for a man'
xhingërrima 'baubles, trifles, trivia'
xhingla-mingla 'trifles, trivia; small ornaments, baubles'
xhixhëlloj 'glitter, glisten'
xhuxh 'dwarf'
xhuxhmaxhuxh 'very short old man [in folklore] with a long beard who lives underground; dwarf'

Moreover, similar facts can be marshaled for the voiced dental affricate <x> in Albanian, as shown in (5.34), suggesting an expressive phonological role for it as well:

(5.34) *xanxar* '(person/animal) with bad habits; mischievous/naughty person'
xarbaxul 'shabbily dressed and dirty person; ragamuffin'[214]
xexerica 'claptrap, nonsense'
xixëlloj 'sparkle, twinkle'

213 The form *xhingla-mingla* is reminiscent of the *m*-reduplication of Greek τζάντζαλα μάντζαλα (see footnote 208), and the *-ma-* of *xhuxhmaxhuxh* may also be related to that process too, but see §4.3.7.2.2 and footnote 262 therein for a different explanation.

214 Note Grk τσαπατσούλης, in (5.32b); the source here is likely to be Trk *çapaçul* 'untidy, slovenly,' where Turkish *ç* is not marked as expressive in and of itself.

xixërimë 'crackling sound (of wood giving off sparks)'
xuq 'shrivelled-up old person who can barely speak; dotard'
xurxull 'soaked from head to toe; stone drunk, soused'
xa 'here you are! '(interjection, Mann 1948: s.v.)[215]

In no case, however, are any of these sounds exclusively found in these expressive contexts; thus, as in Greek, the expressivity of the sounds in question comes largely through the concentration of occurrence they show in expressive lexical items. To the extent, then, that such a distribution offers a suitable glimpse into the functionality of these sounds, a striking parallelism in the various languages emerges.

Another side to expressiveness involving sounds concerns the phonology of onomatopoeia and interjections. Here too there are some noteworthy parallels, though caution is needed due to the potential for universality and thus independent origin in each language. The similarities in question are treated in §4.3.6 in the context of lexical convergence, i.e., (intimate, ERIC-loan) borrowing, but a few reminders here can document the phonological convergence that these expressive lexical items show.

With regard to onomatopoeia, and the like, several languages converge on the representation of the noise made by a dog by showing *h-* as onset and a rime with *au*, *av* or the like; Albanian and Romanian share a somewhat unusual final *-m* in this noise word. To attract a cat, *ps* is widespread, repeated three times (*ps ps ps*). Interjections, as §4.3.4.3 demonstrates, also show convergence, such as Alb *hopa*, Aro *(h)op*, Blg *hop*, Mac *opa*, Grk ώπα, Trk *hop* 'upsy-daisy, oops, etc.' Such lexical items fall in the realm of nonarbitrary connections between form and meaning, and universalist accounts of the similarities cannot be ignored. But where unusual characteristics are evident, like the *-m#* for the Albanian and Romanian dog noise, the parallelism in form is hardly a matter of coincidence or universality alone.

These convergences in various types of phonological expressivity are both striking and interesting, and in the social context of the Balkans, one is indeed naturally led to think about language contact as being involved. A contact-related dimension to all of this is hard to deny, and in the case of interjections and onomatopoeia, simple borrowing is surely at work. As for the expressive functional status of the affricates, that too can be linked to language contact, but probably not as the result of the direct borrowing of special status for the sounds in question. Rather, the involvement of contact is somewhat subtler, though no less real.

One can start with the observation, noted in §5.2 above, that system-external phonology can play an expressive role and can contribute to the "color" and "texture" of an utterance in ways that go beyond the simple denotative function that words have. And indeed, individual sounds can contribute to a "flavor" that a word containing that sound can have. This is true whether the external system is a dialect of the same language or is another language altogether; in either case,

215 See §4.3.4.3.2 on a connection between Albanian *xa* and Greek τζα, used in peek-a-boo (see above (5.31c)).

there are alien sounds for speakers to deal with. Friedman 2002c: 9, for instance, says that schwa in Standard Macedonian imparts a dialectal color to a word and an utterance, since the standard language is based on a western central dialect which happened to have no schwa, whereas just about all of the rest of Macedonian does have schwa, albeit from different sources in different regions. While this involves an indigenous element in the overall Macedonian diasystem, the same could be true of foreign sounds, or foreign sounds that converge on indigenous sounds. This is what seems to be the case with affricates in Greek and in some of the other languages.

It should be clear that many of the words that contribute to the special lexical distribution of the affricates are borrowings, importations into the native system. For instance, Greek and Albanian have τσαπατσούλης (in (5.32c)) and *xarbacul* (in (5.34)) respectively, most likely as borrowings from Trk *çapaçul* 'untidy, slovenly, and Greek has τσατίζω (in (5.31a)) from Trk *çatışmak* 'to quarrel,' and there are others as well. Especially when Turkish words are involved, one has to reckon with a lower stylistic status for the words, as noted by Kazazis 1972 in his discussion of the status of Turkisms in the Balkans in general (and see §4.4). Since lower stylistic status for a word generally means it has a greater degree of colloquialness and is more conversational and informal in nature, and since colloquial language tends to be more expressive and more colorful, less purely denotational and more nuanced with connotation, these borrowings would naturally have fit into the system marked for expressiveness. The sounds contained in such words would thus be prime candidates for taking on an expressive function in and of themselves.

There is moreover another dimension to the role of foreign elements. Expressiveness in a certain sense resides in being out of the ordinary, in being striking and attention-grabbing in some way. Foreign elements by definition are extra-ordinary, literally *exotic*, and typically fall outside of the usual patterns of the recipient language system, of necessity therefore carrying a certain marked and special character to them.[216] In this way, some of the languages which did not play a large role in the Balkan sprachbund, e.g., Italian (see §1.2.2.8), had some impact here in that loans from such languages with affricates would have fit in with, and presumably helped to reinforce, the emerging special character of these sounds in Greek, at least, and possibly elsewhere.[217]

216 Note in this regard footnote 124 in §4.3.1.10, with regard to what Weinreich says about the contribution bilingualism can make to expressivity.

217 Iconicity with the involvement of foreign elements can be pushed further, since among the lexical classes involving expressive τσ in Greek are ones referring to socially marginal groups (cf. (5.32c)) and foreign elements are socially marginal from the start; moreover, in the expressive vocabulary of (5.31) there are ways of interacting with groups at the margins of "normal" (i.e., healthy adult) humankind, namely animals, in (5.31d), and children, in (5.31g). Moreover, even the sound symbolic groups in (5.31a) contain forms referring to edges and points, i.e., limits and boundaries, in a sense marking the limits of the language. See Joseph 1994a for discussion of this more semiotic side to the phenomenon.

While there is undoubtedly more to be said about expressive phonology in the Balkans,[218] much of it is on the lexical level, and all of it shows the classic Balkanistic problem of teasing apart contact effects from those to be attributed to universality and inheritance. Thus while there is certainly something significant in the arena of phonological expressivity to focus on, a precise characterization of the phenomenon and a fuller determination of the role of contact awaits further study.

5.8 Conclusion

By way of conclusion, it should suffice to note that the length of this present chapter exceeds the total number of pages on phonology in the Balkans to be found in the handbooks cited (see §5.1 and footnote 6 above, and also Table 4.1 in §4.1), and that is with much repetition of the same features across the different handbooks. Moreover, this chapter rivals in length at least, the excellent monograph by Sawicka 1997. What this means, therefore, is that there certainly is much to take note of regarding phonology in the Balkans, but the key operative notion is that one has to deal with local Balkan phonologies, not some sort of pan-Balkan phonology.

218 For that matter, expressive phonology in contact situations is seriously under-studied; besides the Balkan material surveyed here, Emeneau 1969 and Householder 1962 discuss one aspect of areal phonic expressivity, namely onomatopoeia, for South Asia and Azerbaijani respectively.