27 Language Contact in the Balkans

BRIAN D. JOSEPH

1 The Languages and Their Convergent Character: Introducing the "Sprachbund"

Southeastern Europe is the home of an intense contact zone that takes in a number of distantly related languages and then some unrelated languages (see p. The Balkans, as this area is generally known, is the most thoroughly studied contact area in the world, and thus occupies a special place in contact linguistics as one of the most important regions for understanding the mechanisms and results of language contact. A rough and mountainous region that forms a peninsula bounded by the Adriatic and Ionian Seas on the west and the Black and Aegean Seas on the east and south, respectively, the Balkans have constituted a crossroads for speakers of many different languages since at least the second millennium BCE.

The interrelations among speakers in the Balkans in ancient times are of considerable interest since clearly various sorts of cross-language transfer showing the effects of language shift (substrata) and borrowing must have occurred. These effects are especially evident in the lexicon – for example, there appears to be a layer of Indo-European but non-Greek words in Ancient Greek (e.g. *aleiphō* 'rub, anoint' where the #*a*- and the -*ph*- are unexpected and the -*lei*- derives from Indo-European **li-p-*, seen in genuine Greek forms such as *lip-os* 'fat') – but other sorts of effects involving various ancient languages could be (and have been) imagined. The languages in question include the following: Continental Celtic (in some form), Dacian (Daco-Mysian), Gothic, Greek, Illyrian, Latin, Macedonian, Phrygian, Pelasgian ("Pre-Greek"), and Thracian. But these pre-historic contacts present a number of challenges to language historians, as some of these languages, quite frustratingly for scholars, are only very sparsely attested or known just from brief mention in ancient sources. And some may not even be identifiable as individual languages.[1] Thus much about their interactions must remain speculative.[2]

For all the intrinsic allure of the study of the ancient pre-historic situation and the speculations about contact that these languages offer, it is the more modern situation, dating from about 1000 CE that has attracted the most attention among researchers in contact linguistics.

The range of languages relevant in the Balkans in this more recent period extends over a number of branches of Indo-European and even beyond that family as well. They include the following (excluding languages, such as Tagalog or Arabic in Greece, that have entered the

Balkans quite recently as the result of employment- or war-related modern immigration trends), listed with some explanatory annotations where deemed appropriate:

(1)   Albanian (both major dialects: Geg (north) and Tosk (south))
       Armenian (spoken in Bulgaria)
       Circassian (Adygey variety; spoken in Kosovo
       Bulgarian
       German (spoken in Romania)
       Greek (including the very divergent dialects like Tsakonian and Pontic, the latter
           only in the Balkans proper via the postwar population exchanges of the 1920s)
       Hungarian (spoken in Romania)
       Italian (spoken in Istria area of Croatia
       Judezmo (also known as Ladino or Judeo-Spanish)
       Macedonian (the Slavic language, thus different from Ancient Macedonian,
           mentioned on p. 000)
       Romani (the Balkan variety of the Indic language of the Roms)
       Romanian (more accurately listed as four separate languages: Aromanian (Vlach),
           Istro-Romanian, Megleno-Romanian, and Romanian (the national language of
           Romania (and Moldova)))
       Ruthenian (also known as Rusyn, spoken in Vojvodina area of Serbia
       "Serbo-Croatian" (now, after the break-up of Yugoslavia, generally considered to be four
           separate ("successor") languages: Bosnian, Croatian, Montenegrin, and Serbian)[3]
       Slovak (in a small enclave in Vojvodina area of Serbia
       Slovenian
       Turkish

Depending in part on just how the Balkans are defined geographically, especially as to the northern border,[4] these languages may be called "languages of the Balkans," a purely geographic designation.[5]

A further distinction, important for contact linguistics, needs to be made here. In particular, one needs to recognize further a class of "Balkan languages," referring to those languages of the Balkans that show considerable structural and lexical convergence due to centuries of intense, intimate, and sustained contact involving multilaterally multilingual speakers. The effect is what one of the first commentators on Balkan convergence, Kopitar (1829: 86), described as an area in which "nur eine Sprachform herrscht, aber mit dreyerley Sprachmaterie" ("only one language-form dominates but with threefold language material"). Under this more restricted designation, the following languages can be included as "Balkan languages," listed again with appropriate annotation as needed for clarification concerning the extent of convergence shown:

(2)   Albanian
       Aromanian
       Bulgarian
       Greek (most dialects, including Tsakonian (but excluding Asia Minor dialects))
       Judezmo (mostly only at the phonological and lexical levels)
       Macedonian
       Megleno-Romanian[6]
       Romani
       Romanian
       Serbian (with Torlak dialects of southeast Serbian being most relevant, much less so
           both the Croatian and Bosnian standards)
       Turkish (not a "full" structural participant but crucial nonetheless).

These Balkan languages represent several different genetic affiliations: Albanian is its own branch within Indo-European, as is Greek; Bulgarian, Macedonian, and (Bosnian-Croatian-Montenegrin-)Serbian are all South Slavic languages, again within Indo-European; the Romanian group and Judezmo belong to the Romance languages (of the Italic branch of Indo-European); Romani belongs to the Indic branch of the Indo-Iranian subgroup of Indo-European; and Turkish is part of the Turkic language grouping, generally believed to be part of a larger Altaic family. It is convenient to refer to these languages more generally as Balkan Albanian, Balkan Greek, Balkan Romance, Balkan Romani, and Balkan Turkish, to distinguish the Balkan varieties of these languages or language groups from their relatives outside the Balkans, inasmuch as the non-Balkan varieties generally do not show the structural and lexical properties that their Balkan relatives do or show them to a lesser degree.

These languages offer some diversity to be sure, but more significantly, also cross-language similarities that have led to the recognition of a key construct for language contact studies – the "sprachbund" (French *union linguistique*, Russian *jazykovoj sojuz*, English *linguistic league* or *linguistic area* or even just *sprachbund*) – that has come to be applied to geographically based convergence zones, including South Asia (Emeneau 1956; Masica 1976), Meso-America (Campbell, Kaufman and Smith-Stark 1986), and the Pacific Northwest of the United States and Canada (Beck 2000), to name just a few. A sprachbund can be defined as any group of languages that due to intense and sustained multilingual contact share linguistic features, largely structural in nature but possibly lexical as well, that are not the result of shared inheritance from a common ancestor nor a matter of independent innovation in each of the languages involved.[7]

Taking note of, and ruling out, common inheritance is an important part of recognizing a sprachbund, inasmuch as this concept, when first formulated by Trubetzkoy in 1923 (with his 1928 pronouncement at the First International Congress of Linguists being the better known and more widely cited source), was explicitly contrasted with a *Sprachfamilie* ('language family'), both being types of *Sprachgruppen* ('language group(ing)s'). In a language family, the languages are genetically related (in the technical linguistic sense of deriving by a direct lineal descent from a common source). Trubetzkoy's new concept, by contrast, was for languages that are geographically related, being located in the same region and often co-existing side by side in the same territory, but are not genetically related, and which yet, due to prolonged contact, show resemblances in form and structure. For Trubetzkoy, the Balkans were the prime example of this new type of grouping. Others before him had made similar observations about the Balkans, though without generalizing to a new construct, let alone offering a handy label for it, or contrasting it clearly with genetic groupings of languages: Kopitar (1829), as noted on p. 000, drew attention to a single feature, the postposed definite article, Miklosich (1861) noted several convergent features involving Balkan languages, among which he included Modern Greek, and Sandfeld (1926), more widely known through the French translation of 1930, elaborated in a systematic way on a large number of such features in phonology, morphology, syntax, and lexicon, with the majority being lexical and phraseological in nature.

2   The Convergent Features Themselves: Balkanisms

It is appropriate at this point to flesh out the somewhat abstract references to structural and lexical convergences with some concrete details. While there are several convergent features – which may be called "Balkanisms" – that have attracted considerable attention in the rather large literature on Balkan linguistics in the period since Sandfeld, appearing for instance in most of the handbooks (Schaller 1975; Banfi 1985; Feuillet 1986, 2012; Asenova 1989/2002; Demiraj 1994/2004; and Steinke and Vraciu 1999) but also in specialized studies

(e.g. Joseph 1983; Friedman 2003), largely because they are widely realized in the Balkan languages, there are actually dozens of features that link small clusters of languages and dialects in the Balkans (see Friedman and Joseph 2020 for details and see also the discussion on p. 000). The following is a representative list of those that are widespread, and a sampling of those that are more localized, thus overall giving a feel for the most significant relevant features and types of features shared by various of these languages; they cover phonology (a–f), morphology (g–j), syntax (k–p), and lexicon (q–r):

(3)   a.   the presence of a (stressed) mid-to-high central vowel; this feature is found in Albanian, Romanian, Bulgarian, some dialects of Macedonian and Bosnian-Croatian-Montenegrin-Serbian, some Romani dialects, and Turkish;

   b.   the presence of *i-e-a-o-u* in the vowel inventory without phonological contrasts in quantity, openness, or nasalization; this feature is found in Greek, Tosk Albanian, Romanian, Macedonian, Bulgarian, Torlak Serbian, and Romani;

   c.   devoicing of word-final stops; this feature is found in Bulgarian, Macedonian, Megleno-Romanian, Modern Greek (dialectally only in one part of northern Greece, as reported in early twentieth century), some Romani dialects, South Montenegrin and Torlak Serbian, and Turkish (somewhat generally but with greater consistency in West Rumelian Turkish);

   d.   development of nasal + voiced stop clusters (e.g. [mb]) out of nasal + voiceless stop combinations, so that the former clusters are rare or nonexistent, or present only in loanwords; this feature is found in Albanian, Aromanian, and Greek;

   e.   presence of ð/θ (voiced/voiceless interdental spirants); this feature is found in Greek, Albanian, Aromanian, and (mostly in loanwords) dialectally in Macedonian;

   f.   realization of /mj/ as [mɲ]; this feature is found in Greek and dialectally in Arvanitika (Tosk Albanian dialects spoken in Greece);

   g.   a reduction in the nominal case system, especially a falling together of genitive and dative cases; this feature is found in Greek, Albanian, Romanian, Bulgarian, and Macedonian (though note that the latter two have eliminated other case distinctions as well);

   h.   the formation of a future tense based on a reduced, often invariant, form of the verb 'want'; this feature is found in Greek, Tosk Albanian, Romanian, Macedonian, Bulgarian, Bosnian-Croatian-Montenegrin-Serbian, and Romani;

   i.   the use of an enclitic (postposed) definite article, typically occurring after the first word in the noun phrase; this feature is found in Albanian, Romanian, Macedonian, Bulgarian, and Torlak Serbian;

   j.   analytic comparative adjective formations; this feature is found in Greek, Albanian, Romanian, Bulgarian, Macedonian, and Romani, as well as in Turkish;

   k.   marking of personal direct objects with a preposition; this feature is found in Aromanian, Megleno-Romanian, and Romanian (via inheritance) and in southern Macedonian dialects;

   l.   double determination in deixis, that is a demonstrative adjective co-occurring with a definite article and a noun (thus, "this-the-man"); this feature is found in Greek, southern Macedonian, and to a limited extent in Albanian and Bulgarian too;

   m.   possessive use of dative enclitic pronouns; this is found in South Slavic and in Greek;

n. the use of verbal forms to distinguish actions on the basis of real or presumed information-source, commonly referred to as marking a witnessed/reported distinction but also including nuances of surprise (admirative) and doubt (dubitative); this feature is found in Albanian, Bulgarian, Macedonian, and Turkish, and to a lesser extent in Romani, Serbian, Aromanian (dialectally), and Romanian (the presumptive);

o. the reduction in use of a nonfinite verbal complement (generally called an "infinitive" in traditional grammar) and its replacement by fully finite complement clauses (see Joseph 1983); this feature is found most intensely in Greek, Macedonian, Bulgarian, Torlak Serbian, and Romani, but also in Albanian (especially Tosk) and Romanian, and to a lesser extent, the successor languages to Serbo-Croatian;

p. the pleonastic use of weak object pronominal forms together with full noun phrase direct or indirect objects ("object doubling"); this feature is found in Greek, Albanian, Romanian, Bulgarian, and Macedonian, dialectally in Serbian, and to a limited extent in Romani;

q. the formation of the "teen" numerals as DIGIT-"on"-TEN; this is found in Albanian, South Slavic, Aromanian, Megleno-Romanian, and Romanian;

r. lexical parallels, including shared phraseology (e.g. a phrase that is literally "without (an)other" meaning "without doubt," or "eat wood" meaning "suffer a beating"), and numerous shared loanwords many of which are from Turkish.

Some of these features are stated as synchronic typological characteristics, e.g. pleonastic use of weak object pronouns, while some are stated in historical terms, e.g. reduction of cases, while still others lend themselves to both sorts of framing, e.g. widespread use of finite complementation due to the replacement of infinitives. Both dimensions – the synchronic and the diachronic – are appropriate to consider in a discussion of the Balkan languages, since it is historical events (of contact and of reaction to that contact) that have led to the convergent typological state found in these languages.

Without belaboring the point, it is important to note that these features are generally taken to be significant indications of contact-induced convergence because, except as noted (e.g. with features involving the various forms of Romanian), they are not features inherited from a common protolanguage (e.g. Proto-Indo-European is typically reconstructed without a definite article and with synthetic (inflectional) analytic adjective formations, so (i, j, l) clearly could not be inheritances, and any features involving Turkish and Indo-European languages similarly could not be due to genetic relatedness). Moreover, it is often the case that they are not found in varieties of the languages outside of the Balkans (e.g. other Romance languages use a preposed definite article and other Slavic languages generally lack an article, and other Romance languages have well-developed infinitival usage, as do other Slavic languages). Further, occasional occurrence of some of these features in other closely or distantly related languages (e.g. a postposed definite article in northern Russian dialects and in Scandinavian languages, or object doubling in Spanish, or a "want"-based future in English) does not vitiate the significant clustering of convergent features in the Balkans. To some extent, then, the geography here allows for what some (e.g. Campbell 1985, 1997: 330–1, 2006: 14) have called a "circumstantialist" argument for a sprachbund. It is also the case that much is known about the history of these languages, and comparisons can be made with earlier stages of Greek, for instance, where infinitival usage abounded, or Slavic, with a well-developed case system, permitting a judgment as to the innovative convergence (and concomitant divergence from earlier structural patterns) and thus allowing for what some (e.g. Campbell 1985, 1997: 330–1, 2006: 14) have called a "historicist" argument for a sprachbund.

3   Causes of Convergence in the Balkans

A key issue in the study of the Balkans is trying to determine what the causes are for the convergences noted. For the most part, scholars agree that language contact is at work, though it must be admitted that some of the developments may well be independent in each language, at least for individual features, even if not for all of the similarities between and among the various languages. Joseph (forthcoming; see also Friedman and Joseph 2020: ch.5) argues that the stressed schwa is not a contact-induced feature but rather one that developed in each language on its own, and aspects of the emergence of the "want"-based future, especially the reductions to a highly "abbreviated" form, may well have taken place on a language-by-language basis, given that full and reduced variants co-existed (or continue to co-exist) in each language for some time.

But even if it is granted that contact is responsible, the question arises as to what kind of contact it was, and what contact-related mechanisms were at work in the formation of the Balkan sprachbund. One can imagine several possibilities:

(4) a. substratum effect (i.e. first-language speakers, shifting to a second language, carry over their habits and structures of the first language into the second, producing an altered form, via interference, of the second language);

   b. adstratum effects (i.e. structures from a second language are imported by speakers into their native (first) language, e.g. for reasons of prestige or communicative ease);

   c. pidginization (i.e. a simplified version of a target language is developed by speakers of several different languages in a situation of communicative necessity);

   d. speaker-to-speaker accommodation to (imperfect) skills of an interlocutor (i.e. a native speaker of the target language adjusts his/her speech to match the perceived level of ability in that language by a non-native speaker; this may involve selection by both speakers of structures for the target language that are "comfortable" to both, that is, acceptable as a variant in the target language and matching some structural element in the other language);

   e. "enhancement via contact" (i.e. a feature which is latent or just emerging in a given language becomes strengthened or increases in frequency due to contact with speakers of another language in which that feature is found).

Several comments about these putative causes are in order. First and foremost, the same feature often has been explained by different scholars in different ways. For example, a substratum explanation has sometimes been proposed for the loss of the infinitive, but the chronology needed to make that work, involving a pre-historic substrate language whose effects surface only in the medieval period, is difficult and argues against it; rather, either an adstratum account (so Sandfeld 1930, with Greek as the prestige language) or pidginization effects (Rozencvejg 1976), or, perhaps better, a mix of pidginization and accommodation drawing on language-internal tendencies can be posited (thus rather like the enhancement scenario of (4e) – see Joseph 1983 for discussion of all of these possibilities).

Second, these causes are not mutually exclusive, in that one might be right for one feature and another right for another feature. The postposed definite article, for instance, unlike the infinitive developments, could well be a real substratum effect, according to the rather compelling account offered by Hamp (1982). But other features are more amenable to other accounts, as the range of possibilities concerning the infinitive-replacement shows.

What all these accounts have in common is that they involve, in one way or another, multilingualism. The question to be asked, then, is what sort of multilingualism is at issue: casual and sporadic or intense and regular, unidirectional or mutual, intimate, or just what?

As the earlier discussion makes clear, it seems that for the Balkans, for the most part (and maybe for sprachbunds in general), what is decisive is intense, intimate, sustained, and mutual multilingualism, involving what Friedman and Joseph (2017, 2020: ch. 8) call their "FOUR-M" model, with multilateral multidirectional mutual multilingualism. Speakers of different languages, living side by side for centuries (which, for the Balkans, corresponds to what is known historically about the co-existence of several languages co-territorially in multilingual villages, towns, and cities), and needing to communicate with one another on a variety of levels, necessarily were familiar with one another's languages to some degree, and accommodated in their usage of their own language to the often imperfect (but possibly quite good) knowledge of that language on the part of speakers of other languages that they interacted with. Speakers of the target language, it can be posited, selected for structures that had ready analogs in the other language, in effect streamlining their own usage in the direction of that of others. Speakers of the other language, for their part, would often have produced structures in the target language that showed the effects of interference (substratum influence) from their own native language. This mutual accommodation on a base of native language interference would naturally lead to the sort of convergence results that characterize a sprachbund. Note that Thomason and Kaufman (1988) posit just such a social context as essential for the development of a sprachbund, namely with the relevant speech communities each maintaining their own linguistic identity in spite of the extensive and intimate contact and thus with some members of the groups of necessity being bi- or multilingual.

Although it might seem that the evidence of overwhelming convergence alone confirms the hypothesis of language contact being involved in the formation of the Balkan sprachbund, there is direct evidence of the sort of contact that breeds a sprachbund, namely the intense and intimate bilingualism referred to . The evidence in question is certain types of lexical borrowings, which requires a bit of explanation. Borrowing of lexical material in and of itself can occur with only casual or even very little contact between speakers; for the latter situation, for instance, the case of learned borrowings through the medium of written texts can be cited. But in some instances, borrowing occurs under circumstances of close and sustained contact in such a way that it is clear that the speakers of different languages must have been communicating with one another on a regular and everyday basis, and under conditions where there was some knowledge of the others' language involved, in short an intimate contact situation with a degree of bilingualism.

The particular borrowings in question are examples of what Friedman and Joseph (2014, 2020: ch.4) have called "ERIC" loans, an acronym for those loans that are "Essentially Rooted In Conversation." They are loans that result from conversational interactions between speakers of different languages in bilingual situations. They are somewhat like the non-need/nonculturally based borrowings that Bloomfield (1933: 461) has called "intimate," but they go beyond that to recognize those that derive from what we may characterize as "human-oriented" interactions among speakers (as opposed to "object-oriented," as with commercial or work-related interactions). They are especially revealing, since they necessarily involve real contact between and among speakers on an intense, regular, and sustained basis, in a more or less equal power situation.[8] Moreover, the ERIC loans to be considered here involve items that are tightly tied to discourse and to the expressiveness that is part and parcel of human discourse. Clearly if such forms pass from language to language, there must have been discourse, i.e. conversational interactions, between speakers of these different languages.

One large area of such discourse-related borrowings involves various sorts of negation. There is no need for borrowing here at all, since the languages of the Balkans – as indeed surely all languages in general – had means for expressing negation. The incorporation of elements of negation from other languages, therefore, must represent the result of close contact among the speakers. Moreover, some of the forms in question have an expressive function that is intimately tied to conversational interaction and is not really found outside of that context, and one, moreover, is paralinguistic and thus could really only spread through visual contact.[9]

For example, Modern Greek and Macedonian have both borrowed the Turkish existential and emphatic negator *yok* 'there is no …; no!' in its emphatic function. Thus Greek has [yok] (spelled <γιοκ>) and Macedonian has *jok*, both with the meaning 'no way; not in the least'.[10] Interestingly, and significantly for the view advocated here, it is the more highly conversationally based function of Turkish *yok* that is borrowed, not the more denotational existential sense. Turkish, for its part, has an interjection *ba* with the meaning 'oh!',[11] which, according to Redhouse (1984), is a borrowing from Greek *ba* (spelled <μπα>) 'ah well' (but also, as a negator, 'unh unh; no way'). And, the widespread Balkan gesture of an upward head nod to signal negation, found at least among speakers of Albanian, Greek, Romanian, and Turkish, may well reflect a diffusion from Greek, given what is known about Ancient Greek gestures and the fact that the distribution especially in Italy coincides with geographic limits of Magna Graecia (Morris et al. 1979); such an element of paralanguage could only spread through face-to-face interaction among speakers, that is, in an intimate conversationally based contact situation.

Other discourse-related borrowings include a large number of interjectional elements, presumably spread through face-to-face contact on a day-to-day basis. For instance, there is a form which can be glossed (roughly) as an "unceremonious term of address" and stems ultimately from Greek (where there are some 55 different variants across Greek dialects – see Joseph 1997), and which, in various of its shapes, is widespread in the Balkans, as indicated in (5):

(5) Turkish: *bre, bire, be*
     Albanian: *ore, or, mor, more, moj, ori, mori, moré, mre, voré, bre*
     Romanian: *bre, mă, mări*
     Bulgarian: *more, mori, bre*
     Macedonian: *more, mori, bre*
     Serbian: *more, mori, bre*
     Greek: *bre, vre, re, are, mare, marí, oré, voré, ori, mbre, pre, more* etc. (this last
                   being the source of practically all these forms)

Similarly, there are several parallel exhortative elements to be found across the Balkans, as given in (6), most probably from Turkish, possibly from *(h)ay* (interjection) + *de* (from *de-mek* 'to say') though there are other possible sources within Turkish:

(6) Romanian: *haide*/(2PL) *haideți*/(1PL) *haidem* 'c'mon; gw'an; let's go'
     Serbian: *hajde*/*hajdemo* (1PL)/*hajdete* (2PL)
     Albanian: *hajde* (SG)/*hajdeni* (PL)
     Greek: *aide* (spelled <άιντε>)

Note in this regard also Bulgarian and Macedonian *ela*, both borrowed from Greek *éla* 'c'mon' (the imperative of 'come').[12] Continuing with interjections, one can also cite Albanian *hopa* and Greek *opa*! 'oops' (for something unexpected), 'woo-hoo!' (expression of joy); Albanian *pa pa pa* and Greek *pa pa pa* 'alas!' (for disgust); and Albanian *aman* and Greek *amán* 'oh my!' (from Turkish *aman*).

Other highly expressive forms that are typically found in colloquial, and thus conversational, usage, also fit in here in that they too have diffused across the Balkans. In particular, one finds in these languages parallels in onomatopoeia (and the like). For example, for a dog's noise, Albanian has *ham-ham*, Romanian has *ham*, Greek has *γav γav*, and Turkish has *hav hav*, and for the noise for attracting a cat, Greek has *ps ps ps*, as also in Bulgarian and Romanian. There is of course the risk of attributing to contact here what might be thought of as universal, but since onomatopes (etc.) do vary across even related and contiguous languages (Spanish has [waw] for the bark of a dog while Portuguese has [kãw]), the Balkan similarities, especially when viewed against the backdrop of other parallels in expressive forms and other grammatical and lexical convergences, fit into a pattern worthy of the attention of contact-minded linguists. In a similar vein, the expressive reduplication with *m-* that is found in Turkish and other more eastern languages,[13] e.g. *kitap-mitap* 'books and such', occurs in Greek, e.g. *dzandzala mandzala* 'this and that', literally 'rags and such' (cf. Levy 1980; Joseph 1984, 1995), Albanian, e.g. *cingrë mingrë* 'trivia', and Bulgarian (cf. Grannes 1978).

Moreover, to elaborate somewhat on the parallel listed on p. 000 in (3r), there are many calques – phraseological loan-translation parallels – in the Balkans in which native language material is substituted for elements in other-language combinations. While there can be learnedisms that are calqued with no direct speaker contact, such as German *Mitleid*, which is a prefix-plus-noun combination based on Latin *compassio* (itself a calque on Greek *sumpatheia* 'compassion, sympathy'), the colloquial and expressive nature of these Balkan calques suggests a social milieu for their creation that is different from that involving learnedisms. In particular, they point to face-to-face speaker interaction. Moreover, they offer direct evidence of bilingualism, since speakers must be familiar enough with the other language to be able to figure out equivalences in their own language to the other language's pieces in the phrase or form or combination of elements being calqued. Thus, as noted in (3r), Greek has *troγo ksilo* for 'I get a beating', but it is literally 'I-eat wood', with the choice of verb agreeing with the Turkish use of *yemek* 'to eat' in the expression *kötek yemek* 'to get a beating' but literally 'a-blow to-eat'. Similarly, what is literally 'to take [someone's] eye' means 'to dazzle' in several languages and 'to cut [one's] mind' means 'to decide'.

Therefore, putting these two types of colloquial and expressive lexical evidence together, a strong case emerges for the conditions being present in the Balkans that were the essential ingredients for the convergence effects that characterize a sprachbund. They thus offer a further argument, along with the geography and the history of the Balkans, that the convergences listed in (3) are indeed indicative of a sprachbund.

4   Assessing the Sprachbund: Localized vs Broadly Realized Convergence

As suggested in the enumeration of Balkanisms given on p. 000 in (3), there are two general types of contact-induced convergences to be recognized:[14] those that occur on a widespread basis among the various languages and those that are highly local in nature. The loss of the infinitive and its replacement by finite forms would be an example of the former type, and the occurrence of prepositional marking of personal direct objects would be an example of the latter type.

From the general approach taken here, with the emphasis on actual speaker-to-speaker contact as the source of the diffusion of features, it should be clear that all diffusion should be taken to be on a localized basis. And, as indicated on p. 000, there are actually dozens more such local convergences to be found in the Balkans.

This observation leads to two questions: first, how one is to reconcile the local effects with the broadly realized Balkanisms, and second, whether, in the face of localized convergences,

it makes any sense to think of the Balkan languages in the broad terms that the "sprachbund" designation requires. In other words, is a sprachbund a viable construct if all the relevant contact takes place locally?

The answer is that local diffusion, if given enough time and the right sort of contact at the relevant "edges" of locales, can lead to spread across larger areas, a sort of "diffusionary chain reaction," as it were. Thus a widely distributed feature, as with the infinitive, would have started small and spread widely from that point, and a feature with a more limited range is one that either has not had the chance yet to spread further or has been checked before spreading further. Indeed, since much remains to be known about the spread of *any* linguistic innovations – those that are the result entirely of language-internal factors as well as those that are contact-induced – it perhaps makes no sense to worry about features that remain localized in their distribution.

What we really have then is *clusters* of convergent languages, where the convergence is on various features in various locales. A sprachbund – the Balkan sprachbund in this case – thus is really to be defined as a cluster of such clusters (see Hamp 1989 especially on this view).[15]

This view of the Balkans, and of sprachbunds in general, has several advantages and addresses a couple of issues typically raised as potential problems for the sprachbund as a viable construct. First, by looking at the overall picture on a feature-by-feature basis, the fact that some convergent features may not be as strongly realized as others is not a problem, since there is nothing that says that all features must be found to the same extent in all languages or even be found to any extent in all of them. Thus, the absence of a postposed definite article from Greek does not vitiate the importance of this feature in linking Albanian, Romanian, and Balkan Slavic. Second, as noted in section 3, different causes may underlie different Balkanisms; they need not be, and probably are not, uniform as to their source. And of the lists of Balkanisms that are generally offered, it is hard to see how all of them must involve contact. Third, the fact that there are differences among the languages even with respect to convergent features – for instance the relatively recent recrudescence of an infinitive in Tosk Albanian (the originally nominal purpose construction of the sort *për të punuar* '(in order) to work' (literally 'for (the-act-of-)working')) in the face of the loss of such constructs in the other languages – is not a problem, since just as features spread on a localized basis, so too can they go off in their own direction on a localized basis.

What this really means is that different features have different histories, but that is as it should be since not all of the features are tied to one another such that a change in one would necessarily trigger a change in another. Moreover, the history is crucial to what the languages are synchronically. In a sense then, in the Balkans, we are dealing with the aftermath of a period of intense contact leading to convergence; the modern standard languages, inasmuch as they were generally formed on the basis of contact-affected dialectal sources, show structural convergences as a relic of their histories; ongoing convergence continues, but on the local level rather than the "national language" level; the conditions that gave rise to the convergence, that created the sprachbund, are no longer present as far as the standard languages are concerned, though they do obtain in various multilingual locales still. Thus in the present just as in the past in the Balkans, the local dialects must be the main focus for the study of language contact, as they are, and have always been, where the action is (see Friedman and Joseph 2017, 2020: ch.8, for more discussion in this regard).

NOTES

1   Eric Hamp has argued, for instance (see Hamp 1994), that "Illyrian" may be a cover term used in ancient times in much the same way that "aboriginal language" is used by many in Australia today

or "Indian language" is used by many in the United States, in each case masking a considerably complex and diverse linguistic situation.

2   The state of what is known about these languages is given a careful treatment in Katičić (1976), and Friedman and Joseph (2020: ch.1) provides a good summary overview.

3   Although these four varieties are similar in many respects, there are some structural and lexical differences holding among them. Moreover, each is a national language: Bosnian of Bosnia and Hercegovina, Croatian of Croatia, Montenegrin of Montenegro (as of 2007), and Serbian of Serbia. Thus, treating them as separate languages is motivated as much by present-day social and political realities as anything purely linguistic.

4   By some accounts, the Balkans may start at Vienna, for instance, while others look to the Danube River as the northern edge.

5   If Tagalog, Arabic, and such languages (including English even) are to be counted, then perhaps one might speak of an even broader designation of "languages *in* the Balkans," of which "languages of the Balkans" would be a subset.

6   Despite distinguishing here among Aromanian, Megleno-Romanian, and Romanian, quite rightly, as separate languages, I nonetheless occasionally, for the sake of convenience, refer simply to "Romanian" as a cover term for all three.

7   There is some controversy as to what the threshold is for recognizing a sprachbund. Thomason (2001: 99) opts for three as this lower limit on the grounds that it trivializes the notion to allow just two languages with structural features in common to determine what should properly be thought of as a special grouping, while Friedman and Joseph (2020: ch.3) argue that convergence is convergence and that therefore, assuming other criteria are met, a two-member sprachbund should not be ruled out in principle.

8   The situation with Romani bilingualism admittedly does not involve equal power structures, since it was unidirectional: Romani speakers learned the other languages around them but speakers of those languages generally did not learn Romani. Still, assuming that what Romani speakers learned was already Balkanized varieties of these other languages, by a process of "reverse interference" (see Friedman and Joseph 2020: ch.3), whereby speaking another language can have an effect on one's native language, Romani speakers could have assimilated their Romani to aspects of these other languages they came to speak.

9   See Joseph (2000, 2001, 2002a, 2002b) for more on parallels in the Balkans involving negation.

10 Turkish also has an emphatic negative, presumably related to *yok*, with the form *yo*. However, despite the similarity to the Albanian word for 'no', *jo*, this Turkish form is unlikely to be the source of the Albanian, since *jo* is found even in the Arbëresh Albanian of southern Italy, an Albanophone area that shows little or no influence from Turkish. I am indebted to Eric Hamp for clarification on this important point.

11 It is of course difficult to give precise definitions for interjections; the glosses here (and further on) are intended just to give a feel for the form's use.

12 Albanian *eja* 'come!' probably has a different origin and is not connected to *ela*; see Joseph (2015) on the etymology of *eja*.

13 And elsewhere – see Southern (2005) on this particular expressive mechanism cross-linguistically.

14 This assumes, of course, that the non-contact-induced convergences, especially those due to separate and independent developments in each language (as is probably the case with stressed schwa, as noted on p. 000), are properly excluded from consideration.

15 Just as a galaxy is made up of constellations and other groupings of stars and planets.

REFERENCES

Asenova, Petja 1989 [2002]. *Balkansko Ezikoznanie: Osnovni Problemi na Balkanskija Ezikov Săjuz* [Balkan linguistics. Basic problems in the Balkan language union]. Sofia: Nauka i izkustvo (2nd edition 2002. Sofia: Faber).

Banfi, Emmanuele 1985. *Linguistica balcanica*. Bologna: Zanichelli.

Beck, David. 2000. Grammatical convergence and the genesis of diversity in the Northwest coast *Sprachbund*. *Anthropological Linguistics* 42: 147–213.

Bloomfield, Leonard 1933. *Language*. New York: Holt, Rinehart, and Winston.

Campbell, Lyle 1985. Areal linguistics and its implications for historical linguistic theory. In Jacek Fisiak (ed.), *Proceedings of the Sixth International Conference on Historical Linguistics*. Amsterdam: Benjamins, pp. 25–56.

Campbell, Lyle 1997. *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press.

Campbell, Lyle 2006. Areal linguistics: a closer scrutiny. In Yaron Matras, April McMahon, and Nigel Vincent (eds.), *Linguistic Areas: Convergence in Historical and Typological Perspective*. New York: Palgrave Macmillan, pp. 1–31.

Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark 1986. Meso-America as a linguistic area. *Language* 62: 530–70.

Demiraj, Shaban 1994 [2004]. *Gjuhësi ballkanike* [Balkan linguistics]. Skopje: Logos-A. (Macedonian edition). 1994: *Balkanska lingvistika* [Balkan linguistics]. Skopje: Logos-A (2nd edition. 2004: Tiranë: Akademia e Shkencave e Republikës së Shqipërisë (Instituti i Gjuhësisë dhe i Letërisë).

Emeneau, Murray 1956. India as a linguistic area. *Language* 32: 3–16.

Feuillet, Jack 1986. *La linguistique balkanique* (Cahiers Balkaniques 10). Paris: INALCO.

Feuillet, Jack 2012. *Grammaire comparée des langages balkaniques*. Paris: Institut d'Études Slaves.

Friedman, Victor A. 2003. Evidentiality in the Balkans. In Alexandra Y. Aikhenvald and Robert M. Dixon (eds.), *Studies in Evidentiality*. Amsterdam: Benjamins, pp. 189–218.

Friedman, Victor A. and Brian D. Joseph 2014. Lessons from Judezmo about the Balkan Sprachbund and contact linguistics. *International Journal of the Sociology of Language* 226: 3–23.

Friedman, Victor A. and Brian D. Joseph 2017. Reassessing Sprachbunds: a view from the Balkans. In Raymond Hickey (ed.), *Handbook of Areal Linguistics*. Cambridge: Cambridge University Press, pp. 55–87.

Friedman, Victor A. and Brian D. Joseph 2020. *The Balkan Languages*. Cambridge: Cambridge University Press.

Grannes, Alf 1978. Le redoublement turk à *m*- initial en bulgare. *Balkansko Ezikoznanie* 21: 37–50.

Hamp, Eric P. 1982. The oldest Albanian syntagma. *Balkansko ezikoznanie* 25.1: 77–9.

Hamp, Eric P. 1989. Yugoslavia: a crossroads of Sprachbünde. *Zeitschrift für Balkanologie* 25.1: 44–7.

Hamp, Eric P. 1994. Albanian. In Ronald E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, pp. 65–7.

Joseph, Brian D. 1983. *The Synchrony and Diachrony of the Balkan Infinitive: A Study in Areal, General, and Historical Linguistics*. Cambridge: Cambridge University Press (reissued in paperback, 2009).

Joseph, Brian D. 1984. Balkan expressive and affective phonology: the case of Greek *ts*/*dz*. In Kot Shangriladze and Erica Townsend (eds.), *Papers for the Fifth Congress of Southeast European Studies (Belgrade, September 1984)*. Columbus, OH: Slavica Publishers (for the US National Committee of the AIESEE), pp. 227–37.

Joseph, Brian D. 1995. Borrowing at the popular level: Balkan interjectional particles of Turkish and Greek origin. *Septième Congres International d'Études du Sud Est Européen: Rapports*. Athens: Greek National Committee for Southeast European Studies, pp. 507–20.

Joseph, Brian D. 1997. Methodological issues in the history of the Balkan lexicon: the case of Greek *vré*/*ré* and its relatives. In Victor Friedman, M. Belyavski-Frank, M. Pisaro, and D. Testen (eds.), *Balkanistica: Studies Dedicated to the Memory of Zbigniew Golab 19 March 1923–24 March 1994*, vol. 10. pp. 255–77.

Joseph, Brian D. 2000. On the development of Modern Greek *óxi* "no." In Chris Schaner-Wolles, John Rennison, and Friedrich Neubarth (eds.), *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of his 60th Birthday*. Turin: Rosenberg and Sellier, pp. 207–14.

Joseph, Brian D. 2001. Language contact and the development of negation in Greek and the Balkans. In *Greek Linguistics '99: Proceedings of the Fourth International Conference on Greek Linguistics, Nicosia, September 1999*. Thessaloniki: University Studio Press, pp. 346–53.

Joseph, Brian D. 2002a. Balkan insights into the syntax of **me:* in Indo-European. In Mark Southern (ed.), *Indo-European Perspectives (Journal of Indo-European Studies Monograph Series 43)*. Washington, DC: Institute for the Study of Man, pp. 103–20.

Joseph, Brian D. 2002b. Language contact and the development of negation in Greek – and how Balkan Slavic helps to illuminate the situation. In Irene Masing-Delic and Mateja Matejic (eds.), *A Festschrift for Leon Twarog: Working Papers in Slavic Studies*, vol. 1. Columbus, OH:

The Ohio State University Department of Slavic and East European Languages and Literatures, pp. 131–9.

Joseph, Brian D. 2015. Inheritance versus borrowing in Albanian etymology: The case of *eja*. In Bardhyl Demiraj (ed.), *Sprache und Kultur der Albaner. Zeitliche und räumliche Dimensionen. Akten der 5. Deutsch-albanischen kulturwissenschaftlichen Tagung* (6.–9. Juni 2014, Buçimas/ Albanien) (Albanologische Forschungen 37). Wiesbaden: Harrassowitz, pp. 196–202.

Joseph, Brian D., forthcoming. *On the Need for History in Balkan Linguistics* (Kenneth E. Naylor Memorial Lecture Series, vol. 10). Oxford, MS: Balkanistica.

Katičić, Radoslav 1976. *Ancient Languages of the Balkans*, 2 vols. (Trends in Linguistics, State-of-the-Art Report 4, 5). The Hague: de Gruyter Mouton.

Kopitar, J. 1829. Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur* 46: 59–106.

Levy, Harry 1980. An Anatolian language-trait in Byzantios' Babylonia and parallel traits on three continents. *MGSA (Modern Greek Studies Association) Bulletin* 12: 47–55.

Masica, Colin 1976. *South Asia as a Linguistic Area*. Chicago: University of Chicago.

Miklosich, Franz 1861. Die slavischen Elemente im Rumunischen. *Denkschriften der kaiserlichen Akademie der Wissenschaften, philosophisch-historische Klasse* 12: 1–70.

Morris, Desmond, Peter Collett, Peter Marsh, and Marie O'Shaugnessy 1979. *Gestures: Their Origins and Distribution*. New York: Stein and Day.

Redhouse, James 1984. *New Redhouse Turkish–English Dictionary*. Istanbul: Redhouse Press.

Rozencvejg, Viktor 1976. *Linguistic Interference and Convergent Change (Janua Linguarum, Series Maior 99)*. The Hague: de Gruyter Mouton.

Sandfeld, Kristian 1926 [1930]. *Linguistique balkanique*. Paris: Klincksieck. (First printed 1926: *Balkanfilologien*. Copenhagen: Luno.)

Schaller, Helmut 1975. *Die Balkansprachen: Eine Einführung in die Balkanphilologie*. Heidelberg: Harrassowitz.

Southern, Mark 2005. *Contagious Couplings: Transmission of Expressives in Yiddish Echo Phrases*. Westport, CT: Praeger.

Steinke, Klaus and A. Vraciu 1999. *Introducere în Lingvistica Balcanică*. Iași: Editura universității "Al. I. Cuza."

Thomason, Sarah G. 2001. *Language Contact*. Washington, DC: Georgetown University Press.

Thomason, Sarah G. and Terrence Kaufman 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

Trubetzkoy Nikolay S. 1923. Vavilonskaja bašnja i smešenie jazykov [The Tower of Babel and the mixture of languages]. *Evrazijskij vremennik* 3: 107–24.

Trubetzkoy, Nikolay S. 1928. *Proposition 16. Actes du Premier congrès international de linguistes*. Leiden: Sijthoff, p. 18.