Reassessing Sprachbunds: A View from the Balkans

By Victor A. Friedman & Brian D. Joseph

ABSTRACT: The Balkan Peninsula in southeastern Europe is home to a number of languages whose interrelationships present various elements of interest for questions of sociolinguistics, language change, language contact, and areal linguistics. The Balkans are the area for which the term "sprachbund" was invented, and with good reason. Of particular linguistic note in the region is the extreme degree of structural convergence — with significant "cultural" lexical commonalities — that many of these languages show. One must start with the key distinction between "languages of the Balkans", a purely geographic designation, and "Balkan languages", encompassing those languages of the Balkans that show considerable convergence. When making such a distinction, the Balkan sprachbund has been classically defined by Albanian, Bulgarian, Macedonian, the Torlak dialects of southern Serbia and southern Kosovo, Greek, Aromanian, Romanian, and Meglenoromanian. The co-territorial Romani dialects, however, are equally part of this construct. Moreover, since the crucial historical period for the formation of the sprachbund as we know it was the Ottoman period, Turkish (especially West Rumelian Turkish dialects), and Balkan Judezmo are participants in the convergence.

Our approach here is first of all to offer some description of the convergence on both the structural and lexical levels. However, with those data before us, we then turn to the social circumstances surrounding and leading to the convergence. Differences in the extent of convergence based on factors such as religion, gender, and social status are brought to light, and our emphasis is on the local conditions in the villages and towns that ultimately aggregate into the entire region. While the Balkans as a whole do indeed constitute a sprachbund in which there is widespread convergence, we argue that it is achieved by intersecting and overlapping micro-zones of convergence, with the mechanisms responsible being bilingualism, code-switching, calquing, and accommodation in face-to-face interactions of a sustained and highly interpersonal level. The sprachbund is thus the result of highly localized convergence in numerous multi-laterally multi-lingual interactive settings, rather than the result of a single process of convergence over a large geographic area. We argue further that a combination of historical record, geography, and social demography offers the best diagnostics for contact-based explanations of the convergences.

Our findings enrich our understanding of the intense language contact situation that creates such zones of historical convergence.

1. Basic facts about sprachbunds, in the Balkans and elsewhere

It is almost impossible to talk about the Balkans from a linguistic standpoint and not utter the term "sprachbund" or one of its less successful English counterparts, such as "linguistic area", "linguistic union", "convergence area", or "linguistic league".[1] Indeed, among linguists, one of the things that the Balkans are best known for is being a sprachbund, that is to say, a zone -- a geographic grouping -- of languages with similarities, especially of a structural nature, that are the result of language contact rather than descent from a common ancestor or typological universals.

The Balkan sprachbund, taking in Albanian, Greek, the South Slavic languages Bulgarian, Macedonian, and some of dialects of the Bosnian-Croatian-Serbian-Montenegrin (BCSM) complex, the Eastern Romance languages Aromanian, Romanian, and Meglenoromanian, the co-territorial dialects of the Indic language Romani, and to some extent the co-territorial dialects of Judezmo (brought to the Ottoman Balkans by Jews expelled from the Iberian peninsula), and Turkish (especially West Rumelian), is noted for a large number of "areal

---

[1] The German *Sprachbund* means literally "language-union"; none of the English terms proposed in its place has really caught on, so we use here the German word as a borrowing into English, and we nativize it. We therefore write it with a lower-case initial letter and form the plural as *sprachbunds*, not *Sprachbünde*; in this way, it is like *pretzel* or other German loanwords in English (plural *pretzels*, not **pretzeln*). As is seen below, just as English *pretzel* has similar but not identical connotations to its German source, so, too, our understanding of sprachbund is not the literal translation from German that has disturbed scholars such as Stolz (2006).

features" — first called "Balkanisms" by Seliščev (1925) — covering aspects of phonology, morphology, syntax, semantics, and lexicon. For concise overviews of these Balkan features, one can consult Friedman 2006a, 2006b, 2011a, Joseph 2003, or Joseph 2010, with more details to be found in handbook-like presentations such as Sandfeld 1930, Schaller 1975, Feuillet 2012, Asenova 2002, Demiraj 2004, and in the compendious Friedman and Joseph 2015.

The result of this linguistic convergence is that, in many instances, one can map between Balkan languages simply by taking note of relevant vocabulary differences.[2] For instance, many of the Balkan languages converge in the use of impersonal (3rd person singular) nonactive verb forms (involving either mediopassive (MP) morphology or a reflexive marker (RX)) with a dative experiencer in the sense 'X feels like...' (literally "to-X$_{DAT}$ VERB$_{NONACT.3SG}$ …"); the convergence is illustrated in (1), with sentences from the various languages with the meaning 'I feel like eating burek':[3]

(1) Macedonian   *mi   se jade   (burek)*
     Aromanian   *nji   -si mãcã   (burec)*
     Meglenoromanian   *ăń   -ţi máncă   (burec)*
                       me.DAT RX eat.3SG.PRS (burek)
     Albanian   *më   hahet   (byrek)*
                       me.DAT eat.3SG.PRS.MP (burek)
     Bulgarian   *jade   mi   se (bjurek)*
                       eats.3SG.PRS me.DAT RX (burek)
     Romani   *hala   pe mange   (bureko)*
                       eat.3SG.PRS RX me.DAT (burek)
     Kastoria Greek   *mi   trójiti   (bureki)*[4]
                       me.ACC   eat.3SG.PRS.MP
                       'I feel like eating (burek)'

Similarly, several of the languages converge in regard to the order of elements marking negation, future tense, mood, and argument structure in the verbal complex, as shown in (2):[5]

(2) Macedonian   ne   ḱe   (da) mu   go   davam[6]
     Albanian   s'   do   të   j+   a   jep
     Romanian   nu   o   să   i+   l   dau
     Dialectal Greek: ðe   θe   na   tu   to   ðóso[7]

---

[2] We adhere to a distinction between the purely geographic designation "language of the Balkans" and the more specifically contact-affected and structurally and lexically convergent "Balkan language"; thus Croatian is a language of the Balkans without being a Balkan language, whereas Macedonian is both.

[3] See Friedman and Joseph 2014a for more on this construction and on impersonals in the Balkans more generally.

[4] This construction does not occur in standard Modern Greek, so the example here is from a northern dialect, one that had been in close contact with Macedonian. The experiencer object is accusative here, reflecting the northern use of the accusative for indirect objects, parallel to the genitive in southern dialects of Greek and datives in other languages.

[5] Albanian and Daco-Romanian here merge dative and accusative weak pronouns into a single portmanteau form; the Greek case is generally labeled "genitive" but it fills typical dative functions.

[6] In Macedonian *da* here is substandard or has a meaning of attenuation. The sign <+> indicates elements written as one word.

[7] We give here dialectal Greek, instead of the standard language, so as to be able to cite an invariant 3rd person form of 'want', *θe*, as the future marker, as it is a more direct parallel with the Macedonian, Albanian, and Romanian markers. The standard language has *θa* (which derives from *θe na*) and while *θa na* (via vowel assimilation) does occur dialectally, *θa* is not the 3rd person form of 'want' so that the parallel would be less direct.

| | NEG | FUT | SBJV | 3SG.DAT/3SG.ACC | give.1SG |

'I will not give it to him.'

This word-by-word or even morpheme-by-morpheme "translatability" between languages is what led Jernej Kopitar (1829:86) to famously characterize the Balkans as an area where "*nur eine Sprachform herrscht, aber mit dreyerley Sprachmaterie.*"[8] Similarly, Miklosich (1861:6-8) remarked on the convergence by noting such features as:

(3) a. future with 'want'+ infinitive[9]
    b. lack of infinitive, with replacement by a finite verb plus a "conjunction"[10]
    c. merger of genitive and dative
    d. the un-Romance postposing of the definite article

It was Trubetzkoy, writing in 1923, who took such observations and coined the notion of *jazykovoj sojuz* 'linguistic union' (cf. *sovetskij sojuz* 'Soviet Union'), whence the German *Sprachbund* in his more famous 1928 formulation (Trubetzkoy 1930).
      In subsequent years, Miklosich's account was significantly expanded and elaborated, esepcially in Sandfeld's 1930 classic, Asenova 2002, and other works, and further features were taken as characteristic of the Balkan sprachbund, as given in (3, continued):

(3) e. replacement of conditional by anterior future
    f. object reduplication (proleptic use of clitic pronouns)
    g. simplification of the declensional system
    h. replacement of synthetic comparatives by analytic
    i. development of a perfect using the auxiliary 'have'
    j. the so-called narrative imperative[11]
    k. evidential forms or usages
    l. certain types of dative subject constructions
    m. shared lexicon from Turkish, Romance, Slavic, Greek, and a presumed ancestor of
       Albanian.

The isoglosses for these and other significant features are complex, e.g., there are various remnants of person marking for some future markers; remnants of the infinitive survive to varying degrees; the conditional meaning of the anterior future extends into the BCSM of Montenegro and Bosnia; the postposed definite article is absent from some regions that have the other features; evidentials sometimes come from perfects and sometimes from futures, and so on

---

[8] "Only one grammar dominates but with three lexicons (literally: 'language material')".
[9] Most of these features were also noted by Kopitar. *Infinitive* here is to be understood only in historical terms. Synchronically, the formulation is 'future marked by a particle descended from 'want' plus finite form (with or without the so-called subjunctive marker)'. Miklosich noted a number of other features, especially phonological ones such as the "prominence" of stressed schwa, and some of these have been repeated in modern surveys. In terms of contact-induced change, however, such features represent parallel historical developments rather than convergence; see Hamp (1977) on the need to distinguish the areal from the typological, and Friedman (2008) on Balkan phonologies versus Balkan phonology (but cf. also Sawicka 1997).
[10] In modern terms, a subjunctive particle. The item in question can also mark single, independent clauses, generally with a modal sense, as in polite commands.
[11] This is a construction wherein an imperative, a form that ostensibly co-indexes a second person subject, can be used with a first or a third person subject, to render narration more vivid.

(see Joseph 1983, Hamp 1989, Friedman 2006b, Greenberg 2000, Belyavski-Frank 2003).[12]

The notion of a geographic area characterized by languages that are similar in various ways owing to contact rather than genealogical heritage has been extended to groupings in other parts of the world, including South Asia (Emeneau 1956, Masica 1976), the Pacific Northwest (Beck 2000), Meso-America (Campbell, Kaufman and Smith-Stark 1986), the Vaupés River region in the Amazon (Aikhenvald 2006b, Epps 2006), the Caucasus (but see Tuite 1999), Ethiopia (Bisang 2006, but see Masica 2001), Mainland Southeast Asia (Enfield 2005), and elsewhere.

In each case that has been studied, there are languages occupying the same space now that have been in that space for a reasonably long time, there are various convergent features, and there is clear evidence of contact among speakers of the different languages in the form of lexical borrowings. Putting that together, especially if there is evidence of divergence on the part of at least some of the languages from an earlier state in which the convergent features are not to be found, one is led to the conclusion that the speaker interactions in the region in question are the reason for the convergence, so that one indeed has a geographic grouping of languages that show structural convergences as a result of contact: a sprachbund.

But, there are reasons to think that identifying a sprachbund is not as simple as it might appear at first glance. In this respect, Tuite (1999) on the Caucasus is highly instructive. He makes the point that while glottalization and certain phraseological calques do indeed appear to be areal in the Caucasus, on careful examination the oft-cited feature of ergativity is in fact realized in such different ways that it cannot be taken as a common element, and that the Caucasus do not, in fact, represent a sprachbund in Trubetzkoy's original sense. In section 2, therefore, we outline some key questions concerning the identification of a sprachbund and then turn to a consideration of how one might answer those questions in general in sections 3 through 6, looking to the Balkans in particular throughout.

2. Interrogating the "sprachbund"

Although it is a well-established concept in the contact linguistics and historical linguistics literature, the notion of "sprachbund" is not without problems. In particular, the following questions are among those that need to be addressed when considering the utility of this construct. These questions pertain to the nature of the languages involved, as stated in (4a), to the nature of the features involved, as given in (4b), and to the nature of the causes, as listed in (4c). Moreover, going beyond these issues, there are others that are also part of the problematization of the notion of "sprachbund". These have to do with the delineation of the region and the groups, as given in (4d), and with the assessment of the construct, as stated in (4e).

(4) Questions to ask concerning "sprachbund" as a viable notion

a. Language-based issues
  i. Is there a minimum number of languages needed before one can identify a sprachbund?
  ii. Must the languages be unrelated to one another? If relatedness is allowed, how closely related can they be?

b. Feature-based issues
  iii. Is there a minimum number of features needed in order to identify a sprachbund?

---

[12] As indicated in note 8, phonological features do not figure significantly in this list. While there are some shared phonological innovations among dialects of different languages at the local level, they are too diverse and diffuse to be generalized in the same way as morpho-syntactic features and the lexicon (Friedman 2008, cf. also Hamp 1977 on schwa as a (non-)Balkanism).

iv. How should the features be distributed across the languages? Must all features be found in all the languages in question? Do some features characterize some languages as forming the "core" of the sprachbund? If so, how does one assess the contribution of the noncore -- peripheral or marginal -- features or languages?

c. Cause-based issues
v. If contact is indeed the basis for the convergence at issue in an area (and not some other causal factor or mere chance), is there a type of contact that is needed in order to identify a sprachbund?
vi. What else might play a role in the formation of a sprachbund?

d. Delineational issues
vii. How do we identify the boundaries of a sprachbund, if any? Are there different degrees of membership as suggested by the core/periphery question in (iv)?

e. Assessement issues
viii. Is the evidence that gives a basis for identifying a sprachbund the effects of past sprachbund construction or is the sprachbund an on-going "organism"?

In the sections that follow, these questions are elaborated upon and some answers given.[13]

3. Answering the Language-based Issues

First, to address (4a.i), since language contact is involved in the basic definition of a sprachbund, clearly the answer to the minimum number of languages necessarily involved is that it is greater than one, but that provokes another question, namely how much greater? Two languages would necessarily constitute the logical minimum for a sprachbund just as the minimum needed in genetic linguistics for a language family — i.e., a situation where no contact is involved — is one, as in the case of so-called *isolates*, languages not demonstrably related to any other.[14] While for a sprachbund, one has to have at least two to tango, i.e. at least some contact between speakers of historically distinct systems, Thomason (2001:99) is among those who insist that a sprachbund must be at least a ménage à trois, a point to which we return below.

Trubetzkoy (1930), in his formulation of the difference between the *Sprachfamilie* and the *Sprachbund*, made no mention of boundaries or numbers. He was attempting both to account for and to distinguish the two diachronic ways languages come to resemble one another, what Labov (2007) has distinguished as *transmission* and *diffusion* (cf. also Hamp 1977). As Trubetzkoy recognized, the *Sprachfamilie* is distinguished by the existence of regular sound correspondences across the member languages, the result of regular sound changes each underwent, that can be determined using the comparative method and can be found in what we can call grammatical morphemes and core vocabulary. The *Sprachbund* on the other hand, as Trubetzkoy defined it, was characterized by shared syntax and morphosyntax, non-systematic phonological correspondences, and common "culture words" (*kulturwörter*).

While the regularity of sound correspondences has a predictability that neatly parallels the scientific method, the distinction between *core vocabulary* and *culture words* is somewhat

---

[13] Much of the material in sections 3 and 4 is adapted from Chapter 3 of Friedman and Joseph 2015, and much of sections 5 through 7 is from Chapter 8.

[14] For example, Sumerian is generally held to be an isolate, as are Zuñi and Basque, to name a few. Such isolates of course can have internal dialect diversity — quite rich in the case of Basque, cf. Trask 1996 — a situation that stretches the notion of "language isolate" through its intersection with the vexing language-versus-dialect question. Also, it may well be that these isolates do in fact form a stock or phylum with some other existing language, but such connections are not demonstrable given our current state of knowledge and methodology.

vague and not immune from social manipulation. Thus, for example, in the Pomak dialects of Greece, numerals and basic kinship terms are Turkish loanwords, despite the fact that the dialects are clearly Slavic in origin, arguably because the speakers of these dialects view Turkish as having importance to their identity as Muslims.[15] Thus, any given body part or basic verb of motion, feeling, bodily function, etc., representing words that would certainly be part of any lexical "core", has the potential to be replaced by a loan.[16] In this respect, Romani is illustrative, perhaps precisely because of its massive multilateral contact. Romani has a pre-Byzantine, mostly Indic, core that accords remarkably well with the notional concept in its basics.

The usefulness of the concept of language family is considered to be self-evident since it provides an historical basis for accounting for language resemblances and relations. At the same time, as noted above, in the absence of evidence demonstrating a relationship of a given language to any other, the existence of a family with only a single member poses no problem to the concept of language family. Similarly, when seeking the defining characteristics of this or that language family, it is precisely the shared history of regular sound change combined with notions of core vocabulary and basic grammar that enable us to speak of boundedness, although Thomason and Kaufman (1988) question the rigidity of such conceptions. It can be argued that nineteenth and early twentieth century ideas connected with the need to establish purities of lineage in "races" was carried over to languages as well, whence Schleicher's (1850:143) characterization of the Balkan languages as resembling one another as being "the most corrupt [*die verdorbensten*] in their families" and Whitney's (1868:199) characterization of structural borrowing as a "monstrosity". In a world suffering anxieties about "purity" of race and origin, and one in which political (national) boundaries were in the process of being drawn and redrawn, it is unsurprising that such concerns would also permeate academic discourse.

Moreover, the difference between a language and a dialect or the definition of a dialect boundary remains, to some extent, a social or political artifact. A particularly telling example from the Balkans is the way in which conflicting Serbian and Bulgarian territorial claims to Macedonia in the nineteenth and first half of the twentieth centuries were bolstered by the selection of this or that isogloss of the South Slavic dialect continuum. Serbian linguists chose the monophonemic (as opposed to diphonemic) reflex of Common Slavic *tj/dj*, while Bulgarian linguists chose the isogloss for the presence of a postposed definite article (see Friedman 2003). Even after the recognition of Macedonian as a separate language within Yugoslavia, Yugoslav linguists continued to treat all dialects with /u/ as the realization of the Common Slavic back nasal as at least "transitional" to Serbian.

In the case of the sprachbund, however, the original point that Trubetzkoy was trying to make is sometimes forgotten or misunderstood. Trubetzkoy was not talking about any situation of bilingual contact but rather situations in which there was a range of similarities in syntax, lexicon, morphosyntax, and even phonology, but precisely without regular sound correspondences and shared core vocabulary. Absent from Trubetzkoy's original formulation but constituting an underlying assumption was areal contiguity, but it is the very nature of areality that raises the question of defining the "area". Masica (1976:11) writes:

---

[15] Likewise both Albanian and Bosnian, with their significant numbers of Muslim speakers, use more kinship terms of Turkish origin than co-territorial or neighboring Macedonian, Serbian, or Croatian. On the other hand, Romani dialects in these regions, whose speakers are predominantly Muslim, have native (or Slavic) kinship terms for these relations, arguably as a boundary-marking device. On the other hand, Bulgarian-speaking Christians from Thrace used Turkish numerals like their Pomak neighbors (Kodov 1935). The point here is that numerals can be highly conservative, but they can also be subject to lexical borrowing.

[16] This includes closed word classes that are generally felt to be resistant to borrowing, such as pronouns or complementizers or conjunctions; see the discussion (and references) in §4 about "ERIC" (conversationally based) loanwords, which defy common assumptions about such resistance.

Some [instances of convergence]... involve only two or three contiguous languages. These may merely be instances of what is possibly a tendency for contiguous languages anywhere in the world—or at least contiguous dialects of contiguous languages—to resemble each other in some way or another. Even if every Indian language turns out to be linked to its neighbors by special two-by-two relationships, forming a continuous network covering the subcontinent, this in itself would not establish India as a special area, especially if similar *arbitrary* [our emphasis VAF/BDJ] linkages continue beyond India....[17]

And Thomason (2001:99) is explicit about numbers:

The general idea is clear enough: a linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor. The reason for requiring three or more languages is that calling two-language contact situations linguistic areas would trivialize the notion of a linguistic area, which would then include all of the world's contact situations except long-distance contacts (via religious language..., etc.)... the linguistic results of contact [among more than two languages] may differ in certain respects.

But, as Hamp (1989) had already pointed out with respect to former Yugoslavia, even the Balkans can be understood as part of a "crossroads of *sprachbunds*", with "a spectrum of differential bindings, a spectrum that extends in different densities across the whole of Europe and beyond". A crucial characteristic of the Balkan sprachbund, and by extension any sprachbund, is that it is, in the words of Thomason and Kaufman (1988:95) "messy," i.e. directionality of the sharing can be difficult to determine. In fact, it is precisely the "messiness" of multiple causation and mutual reinforcement that is characteristic of a sprachbund such as the Balkans. We have enough historical data to know that certain features are innovations in all of the languages for which we have the attestations. While hints may occur in this or that earlier stage of this or that language, and typological parallels may exist elsewhere, in the end we are left with a situation in which languages underwent the same innovation in the same polity at the same time, under known conditions of mutual multilingualism. Under such circumstance, attempting to locate a single cause may not be merely futile, it may in fact smack of the same nineteenth centry anxieties about purity alluded to above. Thus, for example, Leake (1814) attributed the commaonalities of the Balkan languages to Slavic, Kopitar 1829, Miklosich 1861 and others to a substratum of one or more of the unattested or poorly attested ancient languages of the Balkans (by definition unprovable), Sandfeld 1930 to Byzantine Greek, $[try to find one for Latin of maybbe Solta]. Nonetheless given the complexity of the evidence, the attempt to identify a single source for all the Balkan commonalities is demonstrably downright wrong, and the attempt to prove that each innovation arose completely independently in each language strains credulity (see Joseph 1983 on multiple causation, Friedman 2007 on the irrelevance of identifyong a single source for each and every phenomenon, and below on the Janus face of genalogical and areal linguistics). This in turn is because a sprachbund is characteristically a stable situation of mutual multilingualism (cf. also Aikhenvald 2006a), as opposed to, e.g., unidirectional multilingualism such as the vertical multilingualism of the Caucasus (Nichols

---

[17] A problematic aspect of Masica's formulation is the assumption the contiguity entails communication. Such is not necessarily always the case.

1997). A bilingual situation in which the two languages were more or less socially equal and stable would indeed be a type of sprachbund situation.

Thus, for example, if we look to clusterings of smaller, more localized, convergence areas that can be said to add up to a larger area, then, it as indicated above, as few as two languages can be involved in some features that contribute to noteworthy clusterings. For example, the Geg (northern) Albanian variety spoken in the town of Debar in Macedonia matches the local Macedonian dialect in having no vowel nasalization, a phonological detail which occurs in no other northern Albanian dialect. However, other Albanian dialects northeast of Debar have only one nasal vowel, which is a severely reduced inventory for Geg. To this can be added the fact that the peculiar rounded reflex of Common Slavic back nasal matches exactly the reflex of the equivalent vowel in Albanian. To exclude Debar Macedonian and Albanian from consideration in the overall convergence of Macedonian and Albanian, and to treat it as irrelevant to convergences found across other Balkan languages, would be nothing more than an arbitrary decision rooted in acceptance of an aprioristic lower limit in the number of languages needed for a sprachbund.

We are faced then really with two problems: the problem of 'boundaries' (which subsumes both the territorial implication of 'area' and the membership implications of 'union' and 'league') and the problem of 'number'. These problems are reminiscent of the difficulties in defining concepts such as 'nation', 'empire', 'state', 'ethnicity' or 'culture' as well as 'language', 'dialect', 'pidgin', and 'creole'. What level of control constitutes a 'state'? How big must a 'state' be in order to be an 'empire'? How many 'nations' must it comprise? When is the speech of a community a 'dialect' of another 'language' and when is it a separate 'language'? What is 'separate' and how much intelligibility is required before it is 'mutual'?[18] From a general theoretical point of view, it does not actually appear to be the case that the kind of diffusion that takes place among three or more languages is in any way qualitatively different from diffusion that is possible between two languages. For a speaker of a given language -- and the role of speakers in language contact must never be ignored -- diffusion involves the entry of an extra-systemic feature into his/her existing system; whether a feature of system X that comes into system Y was original to system Y or entered Y from another system Z is irrelevant to the reality of the presence of that feature in X and Y (and possibly Z), since in either case, contact and diffusion between X and Y were responsible for the shared presence of that feature. In any case, it is important not to confuse a methodological issue for linguists with the realities of a contact situation for speakers.

Moreover, contact phenomena are never arbitrary. They are embedded in social relations as well as the structures of the languages that manifest them. In a sense, the village of Kupwar in Maharashtra state in India (Gumperz & Wilson 1971, also Masica 1976:11), with its convergence among Marathi, Urdu, and Kannada, is a linguistic area, albeit one that is part of a larger area, just as the dialects spoken in it are parts of larger languages. If Marathi, Urdu, and Kannada can be shown to converge over a broader region at least to some extent, then we can say that in a sense, Kupwar is the Debar of India.

If we keep in mind Trubetzkoy's original motivation in proposing the terminological distinction between *Sprachfamilie* and *Sprachbund*, then two languages related by diffusion can constitute a *Sprachbund* in the same way that two languages related by transmission from an ultimately common source can constitute a *Sprachfamilie*. The crucial difference between the *Sprachfamilie* and the *Sprachbund* is that the former can be an isolate with but a single member, while the latter by definition requires more than one member in order for diffusion to take place.

---

[18] Haugen 1966 remains a classic account of this issue and should be read by those who write about the Balkans regardless of their field.

There is, however, another issue in the definition of a *Sprachbund* as understood by Thomason and Kaufman (1988:95)—who do not impose a tripartite requirement on the concept—namely that of directionality. It is generally agreed that in bilateral language contact situations, there is usually asymmetry in the direction of transference. As argued above, if two languages were demonstrably genetically different enough such that similarities resulting from diffusion could be identified but the directionality could not (or could be shown to be symmetrical) such a unit could arguably be described as a sprachbund. On the other hand, a situation such as that which we find in the Balkans clearly involves diffusion, but directionality can be variable. For our purposes here, determining directionality is desirable but not requisite, and it can even be argued that it is ultimately irrelevant (cf. Ilievski 1973[1988] on the question of internal versus external factors). Moreover, in the end the size question does not affect the Balkans — regardless of the minimum number that might be determined for talking about a sprachbund, it is clear that the Balkans, with up to six distinct contributing language groups (Albanic, Indic, Hellenic, Romance, Slavic, and Turkic), would meet any minimum.

And, as to question (4a.ii) regarding relatedness, it is true that these six contributing language groups constitute only two different language families, Indo-European for the first five and also Turkic.[19] Moreover, within some of these language groups are some very closely related languages, such as Aromanian and Meglenoromanian or Macedonian and Bulgarian. Does this relatedness vitiate the sprachbund in this case? Most assuredly not, if the evident convergence is there and is contact-related, though it does make the job of the linguist somewhat harder, since the essential task of identifying the source of features in a given language suspected of being part of a convergence is complicated by the possibility of the presence of the feature in two related languages being due to (genealogical) inheritance rather than diffusion (contact). Here, as above, though, a methodological issue for linguists, in this case ease of analysis, is not to be equated with realities for speakers.

4. Answering the Feature-based Issues

Points similar to what has been said in §3 can be made in a response to question (4b.iii), regarding the number of features focused on in determining a sprachbund. All researchers here are noncommittal: note Thomason's reference to "*some* [our emphasis, VAF/BDJ] structural features". Clearly, the more convergent features one can find, the more compelling the case becomes for a convergence area, but could a single feature constitute the basis for a sprachbund? Here we can have recourse to Trubetzkoy's original idea, which specified convergences at various levels. There is a parallel to be drawn here with a key aspect of genetic (genealogical) linguistics, namely dialect subgrouping. A single feature that is highly significant, i.e. nontrivial and unlikely to be independently arrived at, can be taken as diagnostic or emblematic for distinguishing one dialect from another and at the same time for establishing subgrouping. The rhoticization of Proto-Albanian *n intervocalically to *r* is perhaps the single most salient feature distinguishing Tosk (southern) dialects of Albanian from Geg (northern), although the isoglosses for the development of stressed schwa from the low nasal vowel, *va* from original *vo* in initial position, and *ua* from *ue* all coincide with the *r/n* isogloss or deviate from it by no more than 20 kilometers.[20] Similarly, the account of Joseph and Wallace 1987 for the parallelism in the form of the first person singular of the verb 'to be' in Latin and Oscan, in that both show a reflex of an

---

[19] Leaving aside here the vexed question of whether Turkic fits into a larger language grouping such as Altaic.

[20] On the level of morphosyntax, there are Tosk dialects with infinitives marked with *me*, usually thought of as a Geg feature. In this case, the *me*-marked infinitives are best viewed as reflecting the proto-Albanian infinitive (so Altimari 2011). This helps support an argument that while regular sound change is genealogically diagnmostic, morphosyntax can be more problematic.

enclitic allomorph of a strong form *esom, can be interpreted as showing that this single innovation is diagnostic of the existence of an Italic branch within Indo-European subsuming these two Indo-European languages of ancient Italy.

Moreover, simply toting up the presence of some set of features and, based on that, scoring the languages as to their degree of membership, as has been done in one form or another various works (e.g. Campbell, Kaufman and Smith-Stark 1986, Reiter 1994, Haspelmath 1998, van der Auwera 1998a, and Lindstedt 2000), fails to provide an accurate picture of a linguistic area, since complex Balkan phenomena are thus treated as unitary, so that that the facts 'on the ground' disappear from view. And what it means to "count" a feature as present in some language is far from a trivial exercise. For instance, the 'feels-like' construction exemplified in (1) is absent from Standard Modern Greek and is found only in some dialects of Greek, specifically in areas where the majority of the population spoke Macedonian, Albanian, and Aromanian until at least the 1920s and in some regions until the end of the 1940s. Does Greek count as a language with this construction? Yes and no. Yes, in that it is found in *some* variety of Greek, but at the same time no, in that it is not widespread across all of Greek. How would that fact, and the fact of the specific dialect distribution of that feature within Greek, be reflected in the scoring? A binary assignment of 1 (presence) or 0 (absence) would not reflect the facts well, nor would a percentage-based score, e.g. 0.1, to signal presence in a small percentage of dialects. And in any case, does the absence across most of Greek make the use of that construction in Kastoria Greek any less significant for the speakers of that dialect? We think not.

Still, even if in principle one should pay attention to a single feature, it is certainly true that the more features that can be conclusively identified as convergent due to contact, the stronger the case is for a sprachbund. Nonetheless, when speaking of features we cannot really identify a quantifiable threshold, a magic number metric for how much shared vocabulary or how many shared features are needed to make a decision regarding relatedness. In this regard, areal linguistics is like its Janus-twin of genealogical linguistics: the criterion for relatedness is simply too many systematic and correspondent similarities to be a coincidence, but attempts to quantify what would constitute coincidence, e.g. Ringe 1992, have been problematic and are not widely recognized as valid.

In the case of the Balkans, for reaching any conclusions about the sprachbund, we also have the advantage of historical documentation for long stretches of time for most of the languages,[21] as well as comparative evidence in the form of related languages, both within the most immediate genealogical groups (e.g. for comparisons with Balkan Romance, not only Italian within Eastern Romance, but beyond that, within Romance more generally, also Spanish and French as well as the non-Balkan dialects of Judezmo) and across more distant relatives (e.g. Celtic and Germanic for comparisons involving the Indo-European branches in the Balkans). Still, while the numerological issue of a minimum can be answered, there is admittedly a qualitative side that cannot be ignored in the assessment of any feature.

Related to the issue of the number of features is question (4b.iv) of what distribution of features is needed in the group to permit classification as a sprachbund. In particular, must the features identified as diagnostic be present in all of the languages? Belić (1936) and Mladenov (1939) adduce the piecemeal geographic distribution of some Balkan features as a problem for the sprachbund construct (so also Birnbaum 1968). The discussion above in §3 addresses that criticism, since a cluster-based approach means that features need not be widespread to be

---

[21] The historical record for Greek begins in ancient (BCE) times, and such is the case as well for Indic (in the form of Sanskrit) and for Romance (in the form of Latin); Turkish is attested from the 8th century and Slavic from the 9th-10th centuries, and Albanian shows up in the historical record in the 15th century.

relevant.[22] Smaller convergence areas that nonetheless overlap, when taken together, determine a more extensive geographic zone in which convergence is to be found. The convergent phonological features seen in Debar Albanian and Macedonian, for instance, when joined with other features that Albanian and Macedonian share, obviously forms more extensive area of convergence involving these two languages. So too, the Kastoria Greek convergence with Albanian and Macedonian and Aromanian regarding the 'feels-like' construction of (1) overlaps with the distribution of the 'want'-based future, found across all of Balkan Greek, and thus adds to the strength of the sprachbund as far as these three languages are concerned.

As with other issues, here, too, a comparison with genealogical linguistics is instructive. According to Bird's (1982) compilation of the distribution of roots reconstructed for Proto-Indo-European in the various branches of that family – Bird operates with 14 such branches – only one root, *tēu- 'swell', is found in all 14 branches. Moreover, there are only eight roots that occur in 13 branches. The number of nonisolated roots increases as the threshold for distributions decreases, so there are 28 roots attested in 12 of the branches, and so on. Taken in this light, the absence of postpositioning for the definite article in Greek, for example, is much less important than its absence in the non-Torlak dialects of the former Serbo-Croatian. Similarly, the distributions of 'have' and 'want' futures take on different significances in different geographic contexts. The point is that it is not the absolute totality of features that all languages share but rather the cumulative effect of smaller convergence zones that justifies the concept of sprachbund.

Having determined that distribution need not be uniform and that the quest for an absolute minimum of determining features is unrealistic without a qualitative assessment of each feature, the next question is whether certain types of features are more relevant than others. The question of the methodological issue of the sprachbund as a consistently definable unit is that despite the parallel first drawn by Trubetzkoy (1923) between the genetic linguistic family defined by common descent and the areal linguistic league defined by subsequent contact, the manner of selecting the correspondences used to define the latter has not been systematized. Contact phenomena, however, do not have the type of systemic invariance found in phenomena such as regular sound change and shared morphology, which serve as the bedrock of demonstrable genealogical origin. Contact-induced change, by its very nature, involves a complex ecology of choices among competing systems (cf. Mufwene 2001). In his original formulation, Trubetzkoy allowed for all types of features (other than those used to define the language family), and, in fact earlier works, such as Miklosich 1861, gave prominence to the lexicon while more recent works give primacy to structural commonalities; Thomason (2001:100), for instance, bases her definition of a sprachbund on "structural features". Within the group of non-lexical features that are taken to be more important, calques are especially significant.[23] Campbell et al. (1986) use evidence from calques and shared metaphors in Meso-American languages to argue for a Meso-American sprachbund, since some degree of bilingualism is needed for calquing to occur and spread (cf. also Ross 2001 on metatypy). Likewise, especially crucial are intimate borrowings and in particular conversationally based loans -- "ERIC" loans (those that are "Essentially Rooted In Conversation") as defined and discussed in Friedman and Joseph 2014b, 2015: Chap. 4 -- taking in various closed classes and

---

[22] Cf. also articles such as Steinke (1999) and Reiter (1999) and the discussion of defining features mentioned in §1 (and in footnote 9). Note also Masica (1976), whose approach to mapping South Asian features involves selecting a few morphosyntactic feature and mapping them in all directions as far as possible.

[23] We call calques "nonlexical" here since they do not necessarily involve the transfer/copying of specific lexical material as opposed to the transfer/copying of conceptual structures; there is often a lexical dimension as well, but it is not requisite (cf. Enfield 2005, 2008 on linguistic epidemiology).

generally borrowing-resistant items including kinship terms, numerals, and pronouns, conversationally based elements such as taboo expressions, idioms, and phraseology, and discourse elements such as connectives and interjections. These are important as they are precisely the lexical items that depend on – and thus demonstrate – close, intimate, and sustained everyday interactions among speakers. They give direct evidence of communication between speakers that is not "object" oriented, not purely towards the end of satisfying needs one speaker may have.[24]

Since contact is involved, except for cases of shared retentions, defining a sprachbund presupposes an innovation and thus a drift not only away from a prior state but toward a state that resembles that occurring in another language; in this way, sprachbund phenomena typically involve both convergence on a new type by two or more languages but concomitant divergence from earlier types (which may be preserved in genetically related languages outside the sprachbund) as well. In numerous oral presentations, Andrei N. Sobolev (U. of St Petersburg, U. Marburg), has claimed that the definition of Balkanisms is circular: Balkan languages have certain features and those features constitute the Balkan sprachbund. This is an unfair, even inaccurate characterization. We begin with the fact that a variety of languages is spoken in a multilingual environment over a long period of time. For most of these languages we have the previous stages well attested. If we leave out this starting point, then it is indeed possible to accuse Balkan linguistics of circularity. However, it is precisely the diachronicity at our disposal that enables us to identify convergent features among the Balkan languages. This is Ilievski's (1973[1988]) point, alluded to above, when he argues that what is important is not the source of convergence but the fact of convergence.

If not all features one is considering need to be found in all the languages one is assessing, then the relevant next question is to query, as in (4b.iv), whether some set of features, and thus some set of languages, constitute the core of the sprachbund, i.e. the most prototypical members as revealed by their incorporation of features x, y, z and so on. This is a reasonable question, and would seem to lend itself to the numerological approaches discussed above, assessing strength of membership by occurrence of features. And in fact, many scholars who have addressed the Balkan sprachbund have talked in terms of different gradations of "membership" in the sprachbund. Schaller 1975, for instance, classifies Balkan languages as "erstes Grades" ('of the first order') and "zweites Grades" ('of the second order'), where some features are taken as more telling than others. Schaller's approach has been seriously criticized ($refs.). Not all features are found to the same extent in all languages; Joseph 1983, for instance, argues that the loss of the infinitive is not realized uniformly across the various Balkan languages, with Macedonian and Greek showing the total loss of the infinitive and the other languages showing infinitives to varying, admittedly limited, degrees.

However, the cluster approach as envisioned by Hamp 1989 and outlined above in §3 provides a basis for understanding the variable realization of given features across the Balkans: each feature has its own spatial trajectory of diffusion, and thus it is unrealistic to expect full "compliance", so to speak, by all languages on all features. The diffusion takes place between speakers, not the languages in some abstract sense, and moreover, it takes place in socially and geographically defined space; this is thus the "speaker-plus-dialect approach" advocated in

---

[24] Other cases where calques have been used in arguing for a sprachbund include Nuckols 2000 with regard to a Central European area taking in Hungarian, Czech, Slovak, and German, and Gil 2011 with regard to a Mekong-Mamberambo linguistic area (covering part of Southeast Asia (the Mekong river area) and down the Malaysia archipelago into western New Guinea (to the Mamberambo river)), where part of the evidence was the shared phraseology of 'EYE-DAY' meaning SUN.

Friedman and Joseph (2015: Chapter 3.3).[25] Further, given that each feature could have a different point of origination, it is evident that some features will have different distributions from others. Thus the cluster approach in a sense means that no particular feature or set of features is privileged as diagnostic for the sprachbund but rather all convergent features contribute to the sprachbund, each in its own locale. The more locales a feature occurs in, the more salient it becomes to the linguist, though not necessarily to the speaker; thus, as in other cases discussed already, the goals of the linguist and the goals of the speaker do not necessarily coincide.

5. Answering the Cause-based Questions

Turning now to the questions pertaining to cause, the response must start with an acknowledgment that speaker-to-speaker contact is responsible for the diffusion of features in the Balkans, as in any sprachbund, and it is therefore responsible for the convergence observed therein. But, as (4c.v) asks, it must be considered whether there is a particular type of contact that is needed. Based on what is seen in the Balkans, the relevant contact is not casual contact but a intense contact, specifically, in the typical case, multi-lateral, multi-directional, mutual multilingualism. This "four-M" model means that several languages are involved (minimally two, based on §3, but in the Balkans and in other cases, more) and that speakers are multilingual, each speaking some version of the language of others, and overall it is mutually so, in that speakers of language X know language Y and speakers of Y know X, with the result that features can flow in either direction from one language to another. The qualifier "some version of" is important because it is not the case that speakers were necessarily perfect bilinguals, rather they had sufficient knowledge of the other language(s) to communicate, and their interlocutors presumably altered aspects of their own usage in the direction of these imperfect speakers.

Moreover, as noted in §4 regarding the sorts of features that are significant, a particular type of loanword is highly significant in dealing with the Balkans, and by extension other sprachbunds. These are the ERIC loans, the conversationally based ones, for they give evidence of the intense and mutual, but also sustained contact that is needed for a sprachbund. The ability of speakers of these different languages to interact on a regular basis in nonneed-based ways, was fostered by a particular socio-historic milieu, and we turn to that by way of answering question (4c.vi), regarding what else, other than contact, might figure in the development of a sprachbund.[26]

Based on our available documentation, processes that may have been set in motion, or at the very least begun to be reinforced during the middle ages, and even features that may have appeared in the written record at earlier dates, achieved their current state during the five-century period of control of the Balkans by the Ottoman Turks, referred to as the *Pax Ottomanica*.[27] Ottoman rule created a socioeconomic stability in the Balkans that allowed for the interactions necessary for the convergence effects that have come to be called by the term "sprachbund". Some comparisons with other contact situations is especially helpful for pinning down what it was about the Balkans that led to the massive convergence that is observed.

Heath (1984:378) presents an interesting view: "It now seems that the extent of borrowing in the Balkans is not especially spectacular; ongoing mixing involving superimposed

---

[25] "Dialect" in the sense of regional or social dialect, distinguished from the standard varieties of the languages in question, which all too often are the basis for comparisons and judgments concerning a sprachbund, at least as far as the Balkans are concerned.

[26] The response that follows draws heavily on Friedman 2014.

[27] In Bulgarian this period is referred to as *turskoto igo* 'the Turkish yoke'. In a similarly telling example of differences of point of view, the Byzantine Empire is sometimes referred to in Bulgarian as *vizantijskoto igo* 'the Byzantine yoke'.

European languages vs. native vernaculars in (former) colonies such as Philippines and Morocco is, overall, at least as extensive as in the Balkan case even when (as in Morocco) the diffusion only began in earnest in the present century". Heath's observations are important for several reasons. On the one hand, the point that significant change can take place rapidly concurs with the view that it was precisely during the Ottoman empire that the Balkan sprachbund as we have come to know it was formed.[28] The examples from Morocco and the Philippines, however, all involve lexical items or reinterpreted morphemes rather than morphosyntactic patterns. Moreover, the relationship of the colonial languages to the indigenous is roughly equivalent to that of Turkish to the Balkan Indo-European languages at the time of the Ottoman conquest. While Turkish did maintain a certain social prestige owing to unequal power relations, there is nonetheless a significant difference between recent European colonial settings lasting a century or so and the five centuries of Turkish settlement in the Balkans during which the language became indigenized and members of all social classes were Turkish speakers. To this we can add that the complexity of indigenous power relations prior to conquest is another part of the picture that is easier to tease out in the Balkans than in European colonies owing to longer histories of documentation of the languages prior to conquest. It is precisely this background of long-term, stable language-contact with significant documentary history that makes the Balkans an interesting model for comparing and contrasting with other contact situations.

It is noted above in §4 that a strictly numerological approach to the Balkans, or any sprachbund for that matter, is fraught with problems. Further consideration of numbers leads to another telling comparison with the Balkans. Hamp (1977) includes a critique of the conflation of areal and typological linguistics seen in Sherzer (1976) in describing indigenous languages of North America. Among Hamp's (1977:282) points is that what he refers to as "gross inventorizing" of what he characterizes as "a Procrustean bed of parameters" (Hamp 1977:283) cannot capture the historical depth and specificity that give meaning to areal developments. Such numbers games played with a small set of features, characterized by Donohue (2012) as "cherry picking", can produce maps in which languages seem to mimic modern politics, e.g. Haspelmath (1998:273), which shows a French-German-Dutch-North Italian "nucleus" to a presumed "Standard Average European," with the Indo-European Balkan languages at the next level of remove, and with Turkish entirely outside of "Europe." A subsequent representation (Haspelmath 2001:107) has only French and German at its core, with Albanian and Romanian as part of the next closest level, Bulgarian and the former Serbo-Croatian beyond that, and Turkish still totally outside. Van der Auwera (1998b:825-827) has dubbed such constructs the "Charlemagne *Sprachbund*" on the undemonstrated assumption that Charlemagne's short-lived (800-814) empire, or its successor the Holy Roman Empire [of the German Nation; a.k.a. the First Reich] was the nucleus for a linguistically unified Europe whose influence can be detected today in mapping out synchronic feature points. This is, in essence, an extension of Sherzer's (1976) methodology to Europe (cf. also König 1998:v-vi), but rather than being the work of a lone researcher, this project — especially in the version known as EUROTYP — has involved many people, produced many volumes, and has taken place in a political context that is arguably motivated by a vision of what Winston Churchill called "a kind of United States of Europe" in his 1946 speech at the University of Zurich. To be sure, as with Sherzer (1976), the assembled data are welcome. The over-arching quasi-historical conclusion, however, is misleading and the lack of attention to historical and dialectological detail of the type called for by Hamp (1977) is problematic.

---

[28] On rapidity of change see Mufwene (2004:203) and Dixon 1997 (regarding "punctuated equilibrium"), but on this latter, cf. also Joseph 2001a.

Van der Auwera's (1998b:827) formulation that on the basis of EUROTYP's investigations "the Balkans do indeed get their *Sprachbund* status confirmed" still gives the impression of treating the Balkan languages like the Balkan states vis-à-vis the EU: their status on the international stage is determined in Brussels (the new Aachen) or Strasbourg (in former Lotharingia).[29] The politics of Western Roman and Eastern Roman (Byzantine) interests, for which the Balkans were always a peripheral but vital pawn, were very much at stake in Charlemagne's time; and the modern-day echoes are striking. But it was precisely the *Pax Ottomanica* of the late medieval and early modern periods — not Obolensky's (1971) Byzantine Commonwealth — in the regions that were part of the Ottoman Empire from the fourteenth to the beginning of the twentieth centuries, where the linguistic realities of the Balkan sprachbund (as identified by Trubetzkoy) took their modern shape. As can be seen from the textual evidence of such innovations as future constructions and infinitive replacement (see Asenova 2002:214, Joseph 2000a), the crucial formative period of the Balkan sprachbund is precisely the Ottoman period, when, as Olivera Jašar-Nasteva said, with one *teskere* (travel document) you could travel the whole peninsula and, we can add, when much of the Charlemagne's former territory consisted of a variety of warring polities that only consolidated into modern nation-states as the Ottoman Empire broke up.[30]

To be sure, Hamp (1977:280) recognizes areal features that "may be crudely labeled Post-Roman European", but, for example, the spread of the perfect in 'have' into the Balkans has nothing to do with Charlemagne. The construction was a Late Latin innovation, whose origins are already apparent in Cicero and Julius Caesar (Allen 1916: 313), and it made its way into the Balkans with the Roman armies, settlers, and Romanized indigenous populations. It became the preterit of choice — independently — in French and Romanian (except in the south; see Pană Dindelegan 2013:33), and continues to displace the aorist in other parts of both Western and Eastern Romance. In Balkan Slavic it was precisely those populations in most intensive contact with the Balkan Romance that became Aromanian that developed independent 'have' perfect paradigms, namely those in what is today the southwest of the Republic of Macedonia and adjacent areas in Greece and Albania (see Gołąb 1976, 1984:134-136 for details). Moreover, it is hardly coincidental that in Bulgarian dialects, it is precisely those that were spoken along the route of the Via Egnatia where similar perfect paradigms developed. As for Greek, as Joseph (2000 and references therein) makes abundantly clear, the use of 'have' as a perfect auxiliary is in fact of very different, albeit also Roman, origin. In Greek, it was the use of 'have' as a future marker — itself a Romance-influenced innovation — that gave rise to an anterior future with the imperfect of 'have' that became a conditional that became a pluperfect that then provided the model for the formation of the perfect using a present of 'have' plus a petrified infinitival form.

---

[29] To a certain extent, this is literally true. In 2005, the European Court of Human Rights in Strasbourg fined the Greek government for violating the human rights of its ethnic Macedonian citizens' in harassing the ethnic Macedonian organization *Vinožito* 'Rainbow'. In 2006, Vinožito used the money to re-publish the 1925 primer that had been published in Athens for Greece's Macedonian minority, combined with a modern Macedonian primer (Vinožito 2006). However, that same year, on September 29, 2006, at the inauguration of Latvian collector Juris Cibuls' exhibition of primers in Thessalonica, the Deputy Mayor for Culture and Youth of that city ordered the organizers to take the Macedonian primer out of the show case so that it could not be displayed (Juris Cibuls, pc).

[30] Differences in territorial, economic, and social mobility are beyond the scope of this paper, but we can note that during the centuries when Jews were locked into ghettoes in Western Europe, and Roms existed there only as peripatetic outcasts, in Southeastern Europe (i.e., Ottoman European Turkey) Roms were settled in both towns and villages (although some groups were peripatetic), and Jews lived in neighborhoods, not locked streets. The larger varieties of available modes of social (and thus linguistic) interaction implied by such differences should not be underestimated. Moreover, pace Haspelmath (2001), significant grammatical change can take place in the course of only a few centuries, as seen in the data in Asenova (2002) and Joseph (2000a); cf. also the changes in English after 1066.

This stands in stark contrast to the Romance perfect, which began as 'have' plus past passive participle, which participle then ceased to agree, which is exactly the construction that was calqued into Macedonian (and some Thracian Bulgarian). On the other hand, the perfect in the Romani dialect of Parakalamos in Epirus (Matras 2004) is clearly a calque on Greek, as is the innovation of a verb meaning 'have'. Albanian also has a perfect in 'have' plus participle, and the participle itself is historically of the past passive type found in Romance and Slavic. The directionality is difficult to judge. The Albanian perfect is securely in place by the time of our first significant texts in the sixteenth century — a time when it was still not well established in Greek — but the relationship to Latin or Romance influence is difficult to tease out. Such perfects are not found in the Torlak dialects of former Serbo-Croatian, a region where there is presumed to have been early contact with populations whose languages are presumed to have been ancestral to Albanian and modern Balkan Romance, and where there were significant Albanian speaking populations until 1878 (Vermeer 1992:107-108). The Slavic dialects of Kosovo and southern Montenegro — where contact with Romance lasted into the twentieth century and with Albanian is on-going, albeit strained — do not show such developments.[31] This fact itself may be due to the importance of social factors in language change. Living cheek by jowl does not necessarily produce shared linguistic structures. A certain level of coexistential communication must also involve social acceptance. On the western end of old Roman Empire, Breton is the only Celtic language with a 'have' perfect, and the directionality is clear. Still to describe all these perfects as part of a "Charlemagne *Sprachbund*" is to do violence to historical facts, although it arguably serves the interest of current political imaginings.

The spread of 'have' perfects exemplifies linguistic epidemiology in Enfield's (2008) sense. And thanks to the depth and detail of our historical records, we can tease out the facts. In some respects, it is heartening to see that humans can program computers to identify, to some extent, insights that humans had without their aid a century or two ago. Thus, for example, as Donohue (2012) demonstrated, WALS (2005) features for the main territorial languages of Europe, when "decoded into binary format, then pushed through computational algorithms (Splitstree) that cluster languages on the basis of 'best shared similarity'" — which he is careful to characterize as explicitly synchronic and not diachronic — produces groupings (moving clockwise from the north) for Germanic, Slavic, Balkan, Romance, and Celtic. The details within the groupings are interesting only because we already know the history: Icelandic and Faroese are closer to German than to Scandinavian, while Afrikaans is closer to Scandinavian than to Dutch, and Polish comes between Belarusan and Ukrainian, on the one hand, and Russian, on the other, while Portuguese is much closer to French than it is to Spanish. Moreover, the ability to differentiate areal from genealogical causality that prompted Trubetzkoy to postulate the sprachbund in the first place, is missing. These results demonstrate clearly Hamp's (1977) point: typological, areal, and genealogical linguistics are independent disciplines, the former achronic, the latter two "twin faces of diachronic linguistics" (Hamp 1977:279). Nonetheless, despite its many sins of omission and commission (under representation of so-called non-territorial languages [itself a problematic, bureaucratic notion], absence of crucial dialect facts, misanalyses, misleading generalizations, etc.), WALS (2005) is a blunt instrument that, if wielded with care and sensitivity, can at least spur us to consider other approaches, as Donohue (2012) has productively done in his discussion of Australia.

---

[31] According to Rexhep Ismajli (p.c.), when Pavle Ivić was conducting field work on the old town former Serbo-Croatian dialect of Prizren (southern Kosovo) in the mid-twentieth century, he gathered a group of old women and asked them to count in the old-fashioned way (*po-starinski*) and they began: *unǎ, dao, trei, patru….* 'one, two three, four' (counting in Aromanian).

In the context of the putative Charlemagne sprachbund, it is instructive to cite here Jakobson's (1931/1971) concept of the Eurasian sprachbund. Jakobson deviated significantly from Trubetzkoy (1930) — who characterized the sprachbund as comprising languages "that display a great similarity with respect to syntax, that show a similarity in the principles of morphological structure, and that offer a large number of common culture words, and often also other similarities in the structure of the sound system" (translation/VAF) — by positing the notion of phonological sprachbunds and specifically a Eurasian sprachbund, concentrating on consonantal timbre (basically palatalization including some correlations with front/back vowel harmony), prosody (presence vs. absence of pitch accent or tone), and, in a footnote, nominal declension. He set up Eurasia as the center in terms of all these. For nominal declension, Germano-Romance Europe and South and Southeast Asia were the peripheries; in terms of phonological tone, the Baltic and Pacific areas were the peripheries (with West South Slavic [most of Serbo-Croatian and Slovenian] as a relic island), while for palatalization the core was roughly the boundaries of the Russian Empire, with the inclusion of eastern Bulgaria (which, perhaps not coincidentally, was imagined as Russia's potential *zadunajskaja gubernaja* 'trans-Danubian province' during the nineteenth and part of the twentieth century). He even went so far as to suggest that palatalization in Great Russian [*sic*] finds its most complete expression, and it is thus no coincidence that Great Russian is the basis of the Russian literary language, i.e. the language with a pan-Eurasian cultural mission (Jakobson 1931/1971:191). All the foregoing is not to say that linguists positing sprachbunds that match political interests intend to act as tools of foreign policy, but once their works are published they can be adopted and adapted by those with policy goals; and in any case, language ideology appears to be at work.

It is also important to note here that, while Masica (2001:239) warns against confusing "recent political configurations" with "linguistic areas," it is precisely the legacy of political configurations such as the Ottoman Empire that created the conditions for the emergence of the Balkan sprachbund as it was identified by Trubetzkoy. At the same time, humans, like all other animals, are capable of traversing whatever barriers nature or other humans might construct, and thus sprachbunds are indeed not political configurations, with fixed boundaries. It is here that the German *Bund* 'union' in *Sprachbund* (as noted in §1, *sojuz* in Trubetzkoy's original 1923 Russian formulation) has misled scholars such as Stolz (2006), who suggests that since sprachbunds do not have clearly definable boundaries like language families (or political entities) the concept should be discarded. His "all or nothing" methodology misses Trubetzkoy's original point that the sprachbund is fundamentally different from a linguistic family, and it fails to take into account the basic historical fact that, like the political boundaries and institutions that sometimes help bring sprachbunds into being, the "boundaries" of a sprachbund are not immutable essences but rather artifacts of on-going multilingual processes (Friedman 2012); in Hamp's (1989:47) words, they are "a spectrum of differential bindings" rather than "compact borders," a point also alluded to in Hamp (1977:282). It is also important to remember that Trubetzkoy first proposed the term at a time when the *Sprachfamilie* 'language family' was widely considered the only legitimate unit of historical linguistics, while resemblances that resulted from the diffusion of contact-induced changes were described in terms such as those used by Schleicher (1850:143), who described Albanian, Balkan Romance, and Balkan Slavic as "agree[ing] only in the fact that they are the most corrupt in their families." Trubetzkoy was explicitly concerned with avoiding the kind of confusion more recently generated by conflations of areal and typological linguistics, although in his time the issues involved areal and genealogical linguistics.

Turning now to language ideology in the Balkans itself, the difference between Greek and the rest is striking. It is certainly the case that multilingualism itself does not guarantee the

formation of a sprachbund. As Ball (2007:7-25) makes clear, in the multilingual Upper Xingu, multilingualism, while necessary for dealing with outsiders, is viewed as polluting, and monolingualism is considered requisite for high status. This endogamous region is quite different from exogamous, parts of Amazonia, where multilingualism is an expected norm, and lexical mixing is viewed negatively, but morphosyntactic convergence is rampant (Aikhenvald 2006b). Consider also the vertical multilingualism that Nichols (1997) has identified as characteristic of the Caucasus, which is similar to various Balkan multilingual practices, where specific types of multilingualism index different types of social status.[32] Ideologies that consider contact-induced change as symptomatic of pollution and that equate isolation and archaism with purity were at work in the nineteenth century as well, as seen in Schleicher's formulation quoted above. We could even suggest that the Charlemagne sprachbund is an attempt both to redress this nineteenth century failing and to co-opt the new valorization of language contact.

6. Answering the Delinealational and Assessment Issues[33]

We are now in a position to address the last of the issues by which the notion of "sprachbund" has been problematized. Question (4d.vii) in a sense restates an issue dealt with earlier, namely that of setting the boundaries of the sprachbund and recognizing different degrees of "participation" in the sprachbund on the part of speakers of the various languages involved. It should be clear from the foregoing discussion what the answer is here: the boundaries are as elastic as the micro-zones of convergence that add up to the larger convergence area. There is nothing fixed, and even political boundaries, while convenient, are relevant only insofar as they correspond to socio-historical realities that might promote the sort of contact necessary for sprachbund formation. Geography is no accident as far as the Balkans are concerned, in that for most of the features recognized as important regarding structural convergence in this region, the more geographically peripheral the language, the less likely it is to demonstrate fully the feature and the more centrally located a language or dialect is in the Balkans, the more fully it shows the feature. This is seen especially clearly with the loss of the infinitive and its replacement by finite subordination. Particularly telling from a geographic standpoint are the comparisons in (5), where [+infinitival] means that the infinitive is alive -- or remained alive longer -- in the language to some (not insignificant) degree and [-infinitival] means that there essentially is no infinitive:[34]

(5)   **[+infinitival]**   **[-infinitival]**
         Romeyka Greek (eastern Turkey)   mainland Greek
         Southern Italy Greek   mainland Greek
         Arbëresh Tosk Albanian (So. Italy)   mainland Tosk Albanian
         Geg   most of Tosk
         West South Slavic (BCSM, Slovene)   East South Slavic (Macedonian/Bulgarian)
         Bulgarian   Macedonian
         Maleshevo-Pirin, Lower Vardar Mac.   the rest of Eastern Mac & all of western
         Croatian   Serbian

---

[32] In vertical multilingualism, people in higher villages know the languages of those down the mountain, but those in the lowlands do not bother to learn highland languages. Nonetheless, as Tuite (1999) makes clear, aside from the features of shared glottalized consonants and a few phraseological calques, when examined closely the idea of a Caucasian sprachbund vanishes like a mirage. Hamp (1977), too, noted that the appearance of glottalization in Armenian, on the one hand, and Ossetian, on the other, must have distinct areal diachronic explanations.
[33] The discussion in this section draws on Friedman and Joseph 2015, especially Chapter 7.7.2 and Chapter 8.
[34] See Friedman and Joseph (2015: Ch. 7.7.2) for details on the various languages that inform this table; most of the relevant facts can be found also in Joseph 1983.

Non-Torlak (N/W) Serbian	Torlak (Southeast) Serbian
Istro-Romanian	Balkan Romance (Aromanian,
	Meglenoromanian, Romanian)
Romanian	Aromanian/Meglenoromanian
East Rumelian & Anatolian Turkish	Western Rumelian Turkish
Modern Indo-Aryan (e.g. Hindi)	Romani
non-Balkan Romani	Balkan Romani

The generalization emerging from such comparisons is that the more deeply embedded a language (or dialect) is in the Balkans, the weaker its category of infinitive is. Admittedly, there are some exceptions to this generalization, e.g. Cypriot Greek is geographically peripheral but lacks the infinitive to the same degree as mainland Greek, and Geg Albanian is relatively central in the Balkans but has an infinitive but Tosk is arguably more central and does not, or at least did not for some stretch of time. While there may be other forces at work in such cases that can help to explain exceptions to the generalization, even with them, it nonetheless holds true in the vast majority of cases and thus provides support for the notion that there is something characteristically Balkan about the loss of the infinitive and its replacement by finite forms.[35]

Peripherality in (5) is geographic in nature, but there can be chronological peripherality as well. Judezmo offers a striking example here. As noted in Friedman and Joseph 2014b, Judezmo entered the Balkans rather late, only after Sephardic Jews were expelled from Spain and Portugal in the late 15th century, and some of the ways in which Judezmo appears to be non-Balkan can be attributed to this chronological dimension. For instance, Judezmo speakers arrived in the Balkans with a fully functioning preposed definite article so absence of the enclitic Balkan definite article can be explained by the fact that an article had already developed in Judzemo (from a Latin starting point without an article); that is, Judzemo was in the Balkans at a point when, in its own development, the issue of a definite article had already been settled. The situation of Judezmo vis-à-vis the infinitival developments is another, perhaps even stronger, case in point. That is, although the ultimate loss of the infinitive in some of the languages is late, or has not yet occurred -- Romanian, for instance, preserves the infinitive as an option even in contemporary usage -- it can be localized temporally in the 16th or 17th century for the languages that lack it most fully, especially Greek, Macedonian, and Tosk Albanian.[36] Thus the relative robustness of the infinitive in Balkan Judezmo, as compared to its linguistic neighbors in the Balkans, especially Greek and Macedonian, may in part be due to chronology. That is, the entry of Spanish Jews into the Balkans, in the late 15th century and afterwards, came at the tail end of the most intense waves of infinitive-loss in the region. One can speculate, then, that being peripheral to the temporal period most associated with strong infinitive-loss may have played a role in the survival of the Judezmo infinitive even into contemporary usage.

Another factor affecting the degree of participation in the convergence builds on the observation in §5 concerning the socio-historical factors. That is, there is also a social dimension to peripherality. Once again, this can be seen with Judezmo and the infinitive. As discussed in Friedman and Joseph 2014b, 2015: Chap. 7.7.2, Balkan Judezmo shows contradictory tendencies regarding the infinitive, with both innovative finite subjunctive usage and conservative infinitival

---

[35] For instance, contact between Greeks of the mainland and Cypriot Greeks -- note that the Cypriots have a separate word for mainlander Greek (*kalamaristika*), suggesting on-going contact -- may have helped to spread infinitivelessness on the Cypriots.

[36] Tosk Albanian has developed a new infinitival construction, e.g. *për të punuar* '(for) to work', apparently composed of a preposition (*për* 'for') with a nominalized participle (*të punuar*), but this seems to be a relatively recent development, after a period of infinitivelessness in the language.

usage. The sociolinguistics of Jewish languages provides a basis for an explanation here. Jewish languages in general are likely to preserve archaisms different from those of coterritorial languages (cf. Wexler 1981), and given the local and social segregation of Jewish communities, Jewish speakers would have less exposure to linguistic innovations found in the usage of coterritorial non-Jewish speakers. The Judeo-Greek of 16th century Constantinople, for instance, shows archaic infinitival usage paralleling that of New Testament Greek (Joseph 2000b). Moreover, as documented in Friedman (1995), Jews were linguistically peripheral as shown by the absence of Judezmo from nineteenth century Macedonian codeswitching anecdotes, where Jews, who speak Turkish, are the only ethnic group that does not switch into its own language. While it is true that non-Jewish merchants in the bazaars often had some knowledge of Judezmo (and, as with Yiddish, Hebraisms were used as cryptolectal elements in such circumstances; Benor 2009), multilingualism, as with Romani, tended to be unidirectional (Friedman 2000). Thus the persistence of the use of infinitives in at least some Balkan Judezmo varieties seems to be an important reflection of a lesser degree of contact between Jews and non-Jews in the Balkans than among the non-Jewish speakers of various languages in the region. By contrast, Romani was in intimate contact with Greek for hundreds of years beginning no later than eleventh century or so. And while we cannot know what proto-Romani looked like at that time, Romani is as dependent as Greek and Macedonian on its analytic subjunctive perform infinitival functions.[37]

Thus peripherality can be measured spatially, temporally, and socially, though social factors of interactions between speakers are the most important. Geography can help or hinder such interactions, and chronology is insurmountable but also instrumental in that speakers can only be in a place when they have come to that place. And, as the lesson of the *Pax Ottomanica* shows, political conditions -- which are a macrosocial phenomenon in any case -- can also serve as a contributing factor. In the end, then, social factors, aided and abetted by various other external conditions, determine particpation and boundaries.

The complexities of Balkan linguistic realities are thus difficult to capture. Still, such difficulty is not to say, with Andriotis and Kourmoulis (1968:30), that the Balkan sprachbund is 'une fiction qui n'est perceptible que de très loin' and that the commonalities are 'tout à fait inorganiques et superficielles'. Their view — which is influenced by Greek nation-state ideology — speaks to the final question, (4e.viii), that of whether the Balkan sprachbund, or any sprachbund that might be identified, is an on-going concern or just the result of processes buried in the past and consignable to the dustbin of history.

Contrary to the challenge posed by Andriotis and Kourmoulis, one can argue Balkan linguistic diversity occurs within the context of a set of structural similarities that constitute a framework of contact-induced change. Moreover, it is important to distinguish between 'superficial' and 'surface'. As Joseph (2001b) argued, surface realizations constitute the locus of language contact, and explanations that appeal to typological aspects of universal grammar (including so-called formalist 'explanations') tell us nothing about language contact.[38] Also, surface realizations are by no means 'inorganic'; they represent convergences that are evidence of the multilingualism that we know existed for centuries and even millennia, and which thrived and developed into the "Four-M" set of conditions under Ottoman rule.

Moreover, and this may be the real source of Andriotis and Kourmoulis's position, when one compares just the standard national languages of the Balkans (as does Asenova 2002, leaving out, however, Macedonian), one can get a sense of a language states frozen in time, fixed

---

[37] Relatively recently, Romani dialects outside the Balkans have begun developing new infinitival constructions in contact with European languages that do have infinitives (Boretzky 1996).
[38] Aikhenvald 2007 makes a similar point.

in a form that might reflect a structural state of affairs from several centuries back, scrubbed somewhat by the process of standardization. Or then again, the literary language might reflect a state of affairs that never existed in the spoken language, but has been borrowed from some prestigious language outside the Balkans. As far as Greek is concerned, for instance, it is instructive to examine the state of the language as described in Thumb 1895 for a time as recent even as the late 19th century and to see forms given as part of regular Demotic usage that exist only in regional dialects today, forced out of the emerging standard language by the onslaught of puristic pressures, overt and covert; for example, the noun for 'thing', currently *práγma* with a genitive *práγmatos*, forms that match Ancient Greek except for the realization of the velar as a fricative, is given in Thumb 1895 as *práma* (still an accepted variant today) but with a genitive *pramátu*, a form found now in such outlying dialects as Cappadocian but not in the standard language. And the effects of lexical purism affecting the ERIC loans are evident in a look through Grannes et al. 2002, where so many of the Turkisms recorded there for Bulgarian are artifacts of the 19th century, considered now, in present-day Bulgarian, to be obsolete or else serving only for offering archaic flavor to literary works. Such is also the case in all the other Balkan languages (Friedman 1996, but see Friedman 2005 on the resurgence of Turkisms in ex-communist countries as a badge of democracy). These examples teach two important lessons.

First, there is the important injunction of Bailey, Maynor, & Cukor-Avila (1989: 299) regarding vernaculars and standard languages:

> ...[T]he history of ... language is the history of vernaculars rather than standard languages. Present-day vernaculars evolved from earlier ones that differed remarkably from present-day textbook[-varieties].... These earlier vernaculars, rather than the standard, clearly must be... the focus of research into the history of... [languages].

This holds true for the Balkans, and the difference between the crystallized structure of the standard languages, with Balkan convergent structure intact but not going anywhere, gives a sense of déjà vu to the sprachbund if one looks only at the standard languages in the Balkans. The standard languages of course are based on speech forms that were affected by the intense contact of the Ottoman period, so that they certainly show Balkanisms. And what are those "present-day vernaculars"? They are the regional dialects, so that the second key lesson is that the dialects and not the standard languages are where the action is; the example in (1) of the 'feels-like' construction in Kastoria Greek but not elsewhere in Greek is a case in point. This means also that on-going contact among speakers of regional varieties of Greek, Macedonian, Balkan Romance, Romani, Albanian, and so on, occurring in northern Greece, in Macedonia, in parts of Albania, in cities and in rural areas, is the present-day analogue to the "Four-M" model conditions (see Friedman 2011b). In that sense, the Balkan sprachbund is alive and continuing to develop as contact among speakers continues.

For sprachbunds that have no standard languages and no literary tradition, the vitality of the convergence zone as a living and on-going entity will depend on the health of the languages involved, i.e. of the speakers' commitment to and ability to use their languages, and the extent to which contact and mutual multilingualism continue. Here the Republic of Macedonia presents a considerably more promising picture than the Hellenic Republic, although both countries are still home to speakers of languages in all the major Balkan linguistic groups. But in principle, a sprachbund is not a relic of the past; under the right conditions, the necessary type of contact renders the processes of convergence on-going.

Thus, while rejecting the notion that the Balkan sprachbund is a fiction or a corpse, it is

crucial to place the differences in the context of the similarities. That is, there are certain cleavages within the Balkans that are particularly revealing, e.g. between 'have'-based futures and 'want'-based futures, or between the presence of systematic marking for evidentiality and the absence of such marking, that yield a crisscross pattern over the whole of the sprachbund area. Paying attention to such features and their distribution offers a more nuanced picture of the sprachbund, one based in part on the recognition that, like all language change, degrees of convergence can take place with varying speeds. And, as with other aspects of the discussion herein, the social conditions hold sway.

7. Conclusion

The problematization and interrogation of the notion of "sprachbund" offered here come down to a single issue: Do sprachbunds exist apart from, i.e. form a entity distinct from, other instances of contact-induced change?

Our answer is yes. That is, first of all, they do exist; there are zones of contact that reflect the effects of intense multi-lateral multi-directional mutual multilingualism. Recognizing such a construct seems to be an inevitable consequence of taking linguistic geography seriously and of studying what is found in key geographic zones linguistically. The sprachbund is a well-instantiated and distinctly observable entity shaped by space and time and by social and political milieu, but at base by speaker-to-speaker contact. This answer, though, leads to additional questions, Are the Balkans an instance of a sprachbund? Again, the answer is yes; whether it is still an instance or instead that the region was once a sprachbund depends on whether one focuses just on comparisons involving the standard languages or instead looks to see the forces that shaped the standard languages structurally and that are still observable in on-going contact situations.

Finally, does a sprachbund represent the outcome of a different *type* of contact? Here the answer is "it depends". It does reflect aspects of contact found in other contexts. e.g. multilingualism, but the type of stable, relatively egalitarian multilingualism may in fact prove diagnostic. The sprachbund, therefore, even if it raises important questions, is a useful construct, a necessary construct even, that contact linguistics needs to recognize within the overall scope of contact-induced linguistic developments.

References

Aikhenvald, Alexandra Y. 2006a. Grammars in Contact: A Cross-Linguistic Perspective. *Grammars in Contact: A Cross Linguistic Typology*, ed. by A. Aikhenvald and R.M.W. Dixon, 1-66. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2006b. Semantics and Pragmatics of Grammatical Relations in the Vaupés Linguistic Area. *Grammars in Contact: A Cross Linguistic Typology*, ed. by A. Aikhenvald and R.M.W. Dixon, 237-266. Oxford: Oxford University Press.

Allen, Joseph H. 1916. *Allen and Greenough's New Latin Grammar*. New York: Ginn & Company.

Altimari, Francesco. 2011. Traces d'infinitifs anciens appartenant à l'albanais dans la langue arbërisht d'Italie. In *L'homme et son environnement dans le Sud-Est européen. Actes de Xe Congrès de l'Association internationale du Sud-Est européen (AIESEE). Paris, 24-26 septembre 2009*, ed. by Hélène Antoniadis-Bibicou, Maurice Aymard, and André Guillou, 445-455. (Textes documents études sur le monde byzantin néohellénique et balkanique 11). Paris: Éditions de l'Association Pierre Belon.

Andriotis, Nikolaos and Georgios Kourmoulis. 1968. Questions de la linguistique balkanique et l'apport de la language grecque. *Actes du Premier Congrès International des Études Balkaniques et Sud-est Européens* 6, ed. by I. Gàlàbov et al., 21-30. Sofia: BAN

Asenova, Petja. 2002. *Balkansko ezikoznanie* [*Balkan linguistics*]. Veliko Tărnovo: Faber [2nd edn.].

Bailey, Guy, Natalie Maynor, and Patricia Cukor-Avila. 1989. Variation in subject-verb concord in early Modern English. *Language Variation & Change* 1.3.285-300.

Ball, Christopher G. 2007. *Out of the Park: Trajectories of Wauja (Xingu Arawak) Language and Culture*. Ph.D. thesis, University of Chicago, Chicago, IL.

Beck, David. 2000. Grammatical Convergence and the Genesis of Diversity in the Northwest coast Sprachbund. *Anthropological Linguistics* 42.147-213

Belić, Aleksandar. 1936. La linguistique balkanique aux congrès internationaux des linguistes. *Revue Internationale des Etudes Balkaniques* 2(=3-4).167-171.

Belyavski-Frank, Masha. 2003. *The Balkan Conditional in South Slavic*. Munich: Otto Sagner.

Benor, Susan. 2009. Lexical Othering in Judezmo: How Ottoman Sephardim Refer to non-Jews. In David M. Bunis, ed., *Languages and Literatures of Sephardic and Oriental Jews*, 65-85. Jerusalem: The Bialik Institute & Misgav Yerushalayim.

Bird, Norman. 1982. *The Distribution of Indo-European Root Morphemes (A Checklist for Philologists)*. Wiesbaden: Otto Harrassowitz.

Birnbaum, Henrik. 1968. Slavjanskie jazyki na Balkanax i ponjatie tak nazyvaemyx jazykovyx sojuzov. *Glossa* 2.1.70-92.

Bisang, Walter. 2006. Linguistic areas, language contact and typology: Some implications from the case of Ethiopia as a linguistic area. *Linguistic areas: Convergence in historical and typological perspective*, ed. by Yaron Matras, April McMahon, and Nigel Vincent, 75-98. London: Palgrave.

Boretzky, Norbert. 1996. The New Infinitive in Romani. *Journal of the Gypsy Lore Society, Fifth Series* 6.1-51.

Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62.530-570.

Demiraj, Shaban. 2004. Gjuhësi ballkanike. Tiranë: Akademia e Shkencave e Republikës së Shqipërise (Instituti i Gjuhësisë dhe i Letërisë) [2nd edn.].

Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Donohue, Mark. 2012. Studying contact without studying the languages involved. Paper presented at the 38th Annual Meeting, Berkeley Linguistic Society, Berkeley, CA, 10-12 February.

Emeneau, Murray, B. 1956. India as a Linguistic Area. *Language* 32.3-16.

Enfield, Nick. 2005. Areal Linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34.181-206.

Enfield, Nick J. 2008. Transmission Biases in Linguistic Epidemiology. *Journal of Language Contact* – Thema 2.297-306.

Epps, Patience. 2006. The Vaupés Melting Pot: Tucanoan Influence on Hup. *Grammars in Contact: A Cross Linguistic Typology*, ed. by A. Aikhenvald and R.F.W. Dixon, 267-289. Oxford: Oxford University Press.

Feuillet, Jack. 2012. *Grammaire comparée des languages balkaniques*. Paris: Institut d'études slaves.

Friedman, Victor A. 1995. Persistence and Change in Ottoman Patterns of Codeswitching in the Republic of Macedonia: Nostalgia, Duress and Language Shift in Contemporary

Southeastern Europe. Summer School: Code-switching and Language Contact, ed. by Durk Gorter et al., 58-67. Ljouwert/Leeuwarden: Fryske Akademy.

Friedman, Victor A. 1996. The Turkish Lexical Element in the Languages of the Republic of Macedonia from the Ottoman Period to Independence. *Zeitschrift für Balkanologie* 32,1.133-150

Friedman, Victor A. 2000. After 170 years of Balkan Linguistics: Whither the Millennium? *Mediterranean Language Review* 12.1-15.

Friedman, Victor A. 2003. Language in Macedonia as an Identity Construction Site. *When Languages Collide: Perspectives on Language Conflict, Language Competition, and Language Coexistence*, ed. by Brian Joseph, Johanna DeStefano, Neil Jacobs, and Ilse Lehiste, 257-295. Columbus: Ohio State University.

Friedman, Victor A. 2005. From Orientalism to Democracy and Back Again. Raymond Detrez and Pieter Plas, eds. *Developing cultural identity in the Balkans: Convergence vs. Divergence*, 25-43. Berlin: Peter Lang.

Friedman, Victor A. 2006a. Balkanizing the Balkan sprachbund: A closer look at grammatical permeability and feature distribution. *Grammars in Contact: A Cross Linguistic Typology*, ed. by A. Aikhenvald and R.M.W. Dixon, 201-219. Oxford: Oxford University Press.

Friedman, Victor A. 2006b. The Balkans as a Linguistic Area. *Elsevier Encyclopedia of Language and Linguistics*, Vol. 1, ed.-in-Chief Keith Brown, 657-672. Oxford: Elsevier.

Friedman, Victor A. 2007. Udvojuvanjeto na objektot vo balkanskite jazici vo minatoto i denes [Macedonian: Object reduplication in Balkan languages in the past and today]. *Klasika - Balkanistika - Paleoslavistika: Festschrift in honor of the 85th birthday and 60th anniversary of scholarly activity of Acad. Petar Ilievski*, ed. by Zuzana Topolińska et al., 235-249. Skopje: Macedonian Academy of Arts and Sciences.

Friedman, Victor A. 2008. Balkan Slavic Dialectology and Balkan Linguistics: Periphery as Center. *American Contributions to the Fourteenth International Congress of Slavists*. ed. by Christina Bethin, 131-148. Bloomington, IN: Slavica.

Friedman, Victor, A. 2011a. The Balkan Languages and Balkan Linguistics. *Annual Review of Anthropology* 40.275-291.

Friedman, Victor, A. 2011b. Le multilinguisme en Republique de Macédoine et l'union balkanique aujourd'hui. In M. Aymard et al., eds., *L'homme et son environnement dans le Sud-Est européen: Actes du Xe Congrès de l'Association internationale des études sud-est européen*, 479-481. Paris: Association Pierre Belon.

Friedman, Victor A. 2012. Conjunction Calquing — A Heartland Balkanism. In T. Kahl, M. Metzelin and H. Schaller, eds., *Balkanismen heute — Balkanisms Today —Balkanizmy segodnja*, 31-37. Vienna: Lit Verlag.

Friedman, Victor A. 2014. Languages are Wealth: The Sprachbund as Linguistic Capital. In Florian Lionnet, Kayla Carpenter, Oana David, Tammy Stark, Vivian Wauters, and Christine Sheil, eds., *Proceedngs: 38th Annual Meeting, Berkeley Linguistics Society*. Berkeley, CA: BLS, University of Calrifornia, Berkeley. <http://elanguage.net/journals/index.php/bls/index>

Friedman, Victor A. and Brian D. Joseph. 2014a. Non-Nominative and Depersonalized Subjects in the Balkans: Areality vs. Genealogy. *Non-Canonically Case-Marked Subjects within and across Languages and Language Families: Stability, Variation and Change*, ed. by Jóhanna Barðdal, Stephen Mark Carey, Thórhallur Eythórsson and Na'ama Pat-El. Amsterdam: Benjamins.

Friedman, Victor A. and Brian D. Joseph. 2014b. Lessons from Judezmo about the Balkan Sprachbund and Contact Linguistics. To appear in *International Journal of the Sociology of*

Language 226.3-23 (Special issue on Jewish Language Contact, ed. by Ghil'ad Zuckermann).

Friedman, Victor A. and Brian D. Joseph. 2015. *The Balkan Languages*. Cambridge, UK: Cambridge University Press.

Gil, David. 2011. Obligatory vs. optional tense-aspect-mood marking. Paper presented at ICHL 20 (20th International Conference on Historical Linguistics), Osaka, July 28.

Gołąb, Zbigniew. 1976. On the Mechanism of Slavic–Rumanian Linguistic Interference in the Balkans. In Thomas Butler, ed., *Bulgaria, Past and Present: Studies in History, Literature, Economics, Music, Sociology, Folklore and Linguistics*, 296–309. Columbus, OH: AAASS.

Gołąb, Zbigniew. 1984. *The Arumanian dialect of Kruševo in SR Macedonia SFR Yugoslavia*. Skopje: MANU.

Grannes, Alf, Kjetil Rå Hauge, Hayriye Süleymenoğlu. 2002. *A dictionary of Turkisms in Bulgarian*. Oslo: Novus.

Greenberg, Robert. 2000. The dialects of Macedonia and Montenegro: Random linguistic parallels or evidence of a sprachbund? *Južnoslovenski filolog* 56.295–300.

Gumperz, J. J. and R. Wilson. 1971. Convergence and Creolization: A case from the Indo-Aryan / Dravidian Border in India. In *Pidginization and Creolization of Languages*, ed. by Dell Hymes, 151-167. Cambridge: Cambridge University Press.

Hamp, Eric P. 1977. On some questions of areal linguistics. In Kenneth Whistler, Robert D. van Valin, Chris Chiarello, Jeri J. Jaeger, Miriam Petruck, Henry Thompson, Ronya Javkin, and Anthony Woodbury, eds., *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, 279–282. Berkeley, CA: Berkeley Linguistics Society.

Hamp, Eric P. 1989. Yugoslavia crossroads of *Sprachbünde*. *Zeitschrift für Balkanologie* 25.1.44-47.

Haspelmath Martin. 1998. How young is standard average European? *Language Sciences* 20.3.271-87

Haspelmath, Martin. 2001. The European Linguistic Area: Standard Average European. In Martin Haspelmath, ed., *Language Typology and Language Universals*, vol. 2, 1492-1551. Berlin: Mouton de Gruyter.

Haugen, Einar. 1966. Semicommunication: The Language Gap in Scandinavia. *Sociological Inquiry* 36.280-97.

Heath, Jeffrey. 1984. Language contact and language change. *Annual Review of Anthropology* 13.367-384.

Ilievski, Petar. 1973[1988]. Kon interpretacijata na modelot na udvoeniot objekt vo makedonskiot jazik. *Godišen zbornik, Filozofski fakultet na Univerzitetot Skopje* 205-220. (reprinted in Ilievski 1988:167-182).

Ilievski, Petar. 1988. *Balkanološki lingvistički studii*. Skopje: Institut za makedonski Jazik.

Jakobson, Roman. 1931/1971. Harakteristike evrazijskogo jazykovogo sojuza. [Characteristics of the Eurasian sprachbund]. *Selected Writings* vol. 1, 143-201. The Hague: Mouton. 2nd ed.

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. 2000a. Textual Authenticity: Evidence from Medieval Greek. In Susan C. Herring, Pieter van Reenen, and Lene Schøsler, eds., *Textual Parameters in Older Languages*, 309-329. Amsterdam: Benjamins.

Joseph, Brian D. 2000b. Processes of Spread for Syntactic Constructions in the Balkans. In C. Tzitzilis & C. Symeonidis, eds., *Balkan Linguistik: Synchronie und Diachronie,* 139-150. Thessaloniki: Aristotle University of Thessaloniki.

Joseph, Brian D. 2001a. Review of R.M.W. Dixon, *The Rise and Fall of Languages*. *Journal of Linguistics* 37.180-186.

Joseph, Brian D. 2001b. Is Balkan Comparative Syntax Possible? In M. L. Rivero & A. Ralli, eds., *Comparative Syntax of Balkan Languages*, 17-43. Oxford: Oxford University Press.

Joseph, Brian D. 2003. The Balkan Languages. In *International Encyclopedia of Linguistics*, Second Edition ed. by William Frawley, 194-196. Oxford: Oxford University Press.

Joseph, Brian D. 2010. Language Contact in the Balkans. In *The Handbook of Language Contact*, ed. by Ray Hickey, 618-633. Oxford: Wiley - Blackwell Publishers.

Joseph, Brian D. and Rex E. Wallace. 1987. Latin *sum* / Oscan **súm**, **sim**, ***esum***. *American Journal of Philology* 108.675-693

Kodov, Hr. 1935 *Ezikăt na trakijskite bălgari* [The language of the Thracian Bulgarians]. Sofia: Trakijskijăt naučen institut.

König, Ekkehard. 1998. General Preface. In Johan van der Auwera, ed., *Adverbial Constructions in the Languages of Europe*, v-vii. Berlin: Mouton de Gruyter.

Kopitar, Jernej. 1829. Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur* 46.59–106.

Labov, William. 2007. Transmission and diffusion. *Language* 83.2.344-387.

Leake, William Martin. 1814. *Researches in Greece*. London: John Booth

Lindstedt, Jouko. 2000. Linguistic Balkanization: Contact-induced change by mutual reinforcement. In Dicky Gilbers, John Nerbonne, and Jos Schaeken (eds.) *Languages in Contact* (Studies in Slavic and General Linguistics 28), 231-246. Amsterdam: Rodopi.

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

Masica, Colin. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In P. Bhaskararao and K. Subbarao, eds., *The Yearbook of South Asian Languages and Linguistics 2001*, 205-268. New Delhi: Sage.

Matras, Yaron. 2004. Romacilikanes: The Romani Dialect of Parakalamos. *Romani Studies 5* 14.1.59-109.

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

Mladenov, Stefan. 1939. Bălgarskijat ezik v svetlinata na balkanistika. *Godišnik na Sofijskija* Universitet, Istor.-filol. fakultet 35.13.1ff.

Mufwene, Salikoko. 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University.

Mufwene, Salikoko. 2004. Language Birth and Death. *Annual Review of Anthropology* 33.201-222.

Nichols, Johanna. 1997. Modeling Ancient Population Structures and Movements in Linguistics. *Annual Review of Anthropology* 26.359-284.

Nuckols, Mark. 2000. *Examination of the case for a Central European Sprachbund of Czech, German, Hungarian and Slovak*. M.A. Thesis, The Ohio State University, Columbus, OH.

Oboblensky, Dimitri. 1971. *The Byzantine Commonwealth*. New York: Praeger.

Pană Dindelegan, Gabriela. 2013. *The Grammar of Romanian*. Oxford: Oxford University Press.

Reiter, Norbert. 1994. *Grundzüge der Balkanologie: Ein Schritt in die Eurolinguistik*. Wiesbaden: Harrassowitz.

Reiter, Norbert. 1999. Von der Balkanologie zur Eurolinguistik. In Uwe Hinrichs, ed., *Handbuch der Südosteuropa–-Linguistik*, 19-26. Wiesbaden: Otto Harrassowitz.

Ringe, Donald A. 1992. On calculating the factor of chance in language comparison. *American Philosophical Society, Transactions* 82.1.1-110.

Ross, Malcolm. 2001. Contact-Induced Language Change in North-west Melanesia, In Alexandra Y. Aikhenvald, and R.M.W. Dixon, eds., *Areal Diffusion and Genetic Inheritance*, 134-166. Oxford: Oxford University Press.

Sandfeld, Kristian. 1930. *Linguistique balkanique*. Paris: Klinksieck.

Sawicka, Irena. 1997. *The Balkan Sprachbund in the Light of Phonetic Features*. Toruń: Energeia.

Schaller, Helmut W. 1975. *Die Balkansprachen. Eine Einführung in die Balkanphilologie*. Heidelberg: Carl Winter.

Schleicher, August. 1850. *Die Sprachen Europas in systematischer Übersicht*. Bonn: König.

Seliščev, A. 1925. Des traits linguistiques communs aux langues balkaniques: Un balkanisme ancien en bulgare. *Révue des études slaves* 5.38-57.

Sherzer, Joel. 1976. *An Areal-Typological Study of American Indian Languages North of Mexico*. Amsterdam: North-Holland.

Steinke, Klaus. 1999. Zur theoretischen Grundlegung der Südosteuropa-Linguistik. In Uwe Hinrichs, ed., *Handbuch der Südosteuropa–-Linguistik*, 67-90. Wiesbaden: Otto Harrassowitz.

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

Thomason, Sarah G. and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California.

Thumb, Albert. 1895. *Handbuch der neugriechischen Volkssprache. Grammatik. Texte. Glossar*. Strasburg: Trübner.

Trask, Robert Lawrence. 1996. *Historical Linguistics*. London: Arnold.

Trubetzkoy, Nikolai S. 1923. Vavilonskaja bašnja i smešenie jazykov. [The tower of Babel and the confusion of languages]. *Evrazijskij vremennik* 3:107-24.

Trubetzkoy, Nikolai S. 1930. Proposition 16. In Cornelis de Boer, Jacobus van Ginneken, Anton G. van Hamel, ed., *Actes du Premier Congrès International des Linguistes à La Haye, du 10–15 Avril 1928*, 17–18. Leiden, The Netherlands: A. W. Sijthoff.

Tuite, Kevin. 1999. The Myth of the Caucasian Sprachbund: The Case of Ergativity. *Lingua* 108.1-26.

Van der Auwera, Johan. 1998a. Revisiting the Balkan and Meso-American Linguistic Areas. *Language Sciences* 20.3.259-270.

Van der Auwera, Johan. 1998b. Conclusions. In Johan van der Auwera, ed., *Adverbial Constructions in the Languages of Europe*, 813-836. Berlin: Mouton de Gruyter.

Vermeer, Willem. 1992. Serbs and Albanians in Yugoslavia. *Yearbook of European Studies* 5.101-124.

Vinožito 2006. *Bukvar/Anagnōstiko* ['Primer' Macedonian/Greek]. Thessaloniki: Batavia.

WALS. 2005. *The World Atlas of Linguistic Structures*, ed. by Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press.

Wexler, Paul. 1981. Jewish interlinguistics: Facts and conceptual framework. *Language* 57.99–149.

Whitney, William D. 1868. *Language and the Study of Language: Twelve lectures on the principles of linguistic science*. New York: C. Scribner & Co.