Abstract

The relationship between areal linguistics and typology can be illuminated and clarified by Macedonian dialectology, especially in the context of Balkan linguistics. This is especially important in analyses of the Balkan linguistic league, which, as an areal phenomenon, should be seen in historical perspective. On the one hand, Macedonian dialectology helps clarify the graded nature of the phenomenon itself, and on the other helps demonstrate that Macedonian is the most Balkanized of the Balkan languages.

In this paper I wish to examine the importance of Macedonian dialectology for a more precise understanding of the Balkan linguistic league in general, and in the context of modern typological linguistics in particular. I shall focus on the dialectological facts concerning a number of commonly cited and some less commonly cited Balkanisms, i.e. contact-induced shared innovations among the languages of the Balkan linguistic area. In examining Balkan linguistic phenomena, both feature selection and degree of grammaticalization are crucial factors. I shall conclude that Macedonian is indeed the most Balkanized of the Balkan languages, and that this should be seen in historical perspective.

The study of Macedonian has contributed much to Balkan linguistics, especially since the formation of the field as such in the basic works of Šeliščev (1918, 1925) and Sandfeld (1926/1930). Thus, for example, Šeliščev (1918: 250-259) distinguishes the western Macedonian dialects as the most Balkanized, e.g., in connection with object reduplication. Likewise, Sandfeld (1926/1930), despite the fact that he worked at a time during which the label ‘Bulgarian’ was usually applied to both the Bulgarian and the Macedonian languages, nonetheless distinguishes macëdo-bulgare and bulgare with great consistency. From this we can infer that Sandfeld was conscious of the special place that Macedonian dialects occupy within the Balkan linguistic league, and in this context it is interesting to note that Sandfeld only rarely specifies his Albanian examples as Geg or Tосk. It can therefore be argued that Sandfeld treats the dialects of Macedonian as a distinct unit, thus preceding by a number of years Vailant’s (1938) work, in which he argues in detail for the need to accord Macedonian a special place within South Slavic. In recent decades, the distinctness of Macedonian is no longer a subject for debate except in some retrograde Bulgarian linguistic circles (e.g. Kočev 2001, see Friedman 2004), and thus recent typological studies generally treat Macedonian as a distinct linguistic unit.

On the other hand, European typological linguistics, or Eurolinguistics, which can be seen as part of a larger field of European Studies, or Eurology, has assigned a peripheral place to the Balkans in the typological linguistic map of Europe (e.g., Haspelmath 1998:273, where the center is placed at the Romance-Germanic border), and within the Balkans it has been suggested that Bulgarian occupies a
more central place than Macedonian (Van der Auwera 1998). Most recently, Heine and Kuteva (2006) have suggested that all the languages of Europe are approaching a common type, and that this convergence dates from the Roman Empire. This thesis is reminiscent of Sandfeld’s insistence on the Byzantine Empire as the source of the Balkan linguistic league, but with a westward shift of the center of gravity more consistent with scholars who see the Roman occupation of the Balkans as responsible for introducing either specific Balkan traits or the conditions of multilingualism that led to the development of those traits (see Solta 1980; see also Golab 1960, 1984 for a balanced view of the role of Romance in the Balkans).

Before turning to the examination of specific traits (Balkanisms in the linguistic sense of the term), we must give a brief discussion of the identification of these traits in general. On the one hand, the linguistic features that characterize the Balkan linguistic league as such can be descriptively labeled, those labels listed, and the lists themselves quantified. In fact, without some sort of taxonomy, it would be impossible to discuss Balkan linguistic phenomena in a meaningful or insightful way. At the same time, however, the lists and their quantifications must be carefully nuanced, and, moreover, must be kept in semiotic perspective: indices of complex linguistic phenomena can refer to but not replace the analyses on which they are based. A number of recent works on Balkan linguistics have attempted to synthesize our understanding of the Balkan linguistic league via lists and analyses based on lists, and among these Van der Auwera (1998) can be taken as an interesting example especially worthy of discussion from the point of view of modern linguistic typology. I shall therefore take this article as the basic frame for my discussion.

The list in question is taken from Joseph (1992) with the addition of one feature and can be summarized as follows: A) Phonological a) stressed schwa; b) five-vowel system without distinctive length, openness, or nasalization; B) Morphological a) genitive-dative merger; b) future tense with reduced form of ‘will’; c) postponed definite article; d) reported, admiration, and dubitative verb forms [and usages]; C) Syntactic a) infinitive loss; b) analytic adjectival gradation; c) object doubling; d) teens formed by ‘numeral-on-ten’ (this last being the added feature). The methodology is to count the relative presence or absence of a feature in a given Balkan language or language of the Balkans, assign it numerical value, and then map the relative degrees of Balkanization by means of isopleths rather like the feature maps in Masica (1976). In terms of the selection of languages and dialects, the study, being based on Joseph (1992), makes use of standard Bulgarian, Macedonian, Romanian, Greek, Turkish, Bosnian/Croatian/Serbian (BCS, the former Serbo-Croatian, especially the Torlak dialects of the southern regions), as well as Geg and Tosk Albanian, and Romani (dialects unspecified).
Mentioned by Joseph (1992) but absent from van der Auwerda (1998) are Aromanian, Megleno-Romanian, and Judezmo, while Gagauz is left out of both.¹

To begin with Balkan phonology (for a detailed treatment of this topic, see Sawicka 2001), it can be argued that it is precisely in phonology that the Balkan languages generally preserve striking differentiating specificities at the macro-level, although particular developments can be locally shared (see HAMP 1977). It is thus the case that when the speaker of one Balkan language speaks another, the phonology will often be based on that of the first language. The fact that some people are described as speaking ‘without an accent’ is the exception which proves the rule.

As just indicated, however, this is not to say that there are no phonological adaptations at the dialectal level. Thus, for example, Aromanian and some Macedonian dialects in contact with Greek in present-day Greek Macedonia have inter-dental fricatives in Greek loanwords and even in some non-Greek words; Romani dialects in contact with Greek have /s/ for earlier /š/; various Slavic and Romani dialects in intensive contact with Turkish and/or Albanian have /ü/, while some Turkish dialects in intensive contact with Albanian and Macedonian (and Torlak BCS) lose /ö/. A significant diachronic development in the Debar town dialects of Macedonian and Albanian is shared reflex of the low back nasal of Common Slavic and Common Albanian as the same denazalized open /o/, which contrasts with a more closed /ö/. The loss of /h/ in the western Balkans can certainly be viewed from an areal perspective as the result of shared linguistic habits; however, the behavior of laterals presents a much more complex picture in Albanian vis-à-vis Serbian (see HAMP 2002, who posits Albanian speech habits influencing the Serbian outcome), in northern versus southern dialects of Macedonian (where the north resembles Serbian), and also in northern versus southern dialects of Greek (where the north resembles Macedonian and Bulgarian), as well as in the Turkish and Romani dialects of the Balkans (which also tend to use Balkan Slavic rules of clear /l/ before front vowels and dark /l/ elsewhere).²

Thus when speaking of phonology in the context of Balkan linguistics, it is more accurate to speak of Balkan phonologies rather than Balkan phonology.

With regard to stressed schwa as a Balkanism (first proposed in Miklosich 1861), Macedonian does in fact have the vowel in the majority of its dialects (albeit from different sources in different regions), and moreover, the development was universally present during the late medieval and early modern periods. Thus, the first problem with a typological approach to areal linguistics (as already

---

¹ The omission of Romani, Judezmo, and Turkic dates from Sandfeld’s (1926/1930) explicit exclusion of them, although more recent work has demonstrated that the Balkan dialects of these languages do participate significantly in Balkan linguistic processes. See Friedman 2000 for discussion and references.

² On the other hand, we also have clear /l/ before back vowels in Macedonian in certain loans from Turkish, where Turkish itself has borrowed the words from Arabic or Persian.
indicated by HAMP [1977:281]), and one which is brought out by Macedonian dialectology, is that crucial details and historical facts can be missed.

In connection with the other developments, between the level of morphology and the level of syntax there is the morphosyntactic zone, which is where the most typical (versus typological) Balkan linguistic phenomena occur. A significant feature omitted from the abovementioned list of Balkanism that is one of the most important is the Balkan conditional, formed by the intersection of future and imperfect markers (GOLAB 1964, BELYAVSKI-FRANK 2003). Not only is this type of conditional typical of Balkan Slavic, Greek, Albanian, and Balkan Romance south of the Danube, but its complete grammaticalization in Macedonian (especially the west and southwest), Aromanian, Megleno-Romanian, Tosk Albanian and Greek—versus the lesser degree of grammaticalization in Bulgarian, Romanian and Geg Albanian—is an important part of the Balkan linguistic picture.

On the other hand, with regard to the ‘on-ten’ method of forming teens, HAMP (1992) argues convincingly on the basis of gender differentiation (in Albanian and Balkan Romance ‘ten’ is feminine whereas in Slavic [as in Latin and the rest of Romance] it is masculine)) that the ‘on-ten’ innovation of forming teens occurred when pre-Albanian and proto-Slavic were in contact in northern Europe (which contact is supported by shared ancient loans, developments of Winter’s law, etc., see HAMP 1994). The speakers of the language that became Latinized into Romanian thus have inherited a pre-Albanian pattern rather than borrowed a Slavic one. As Hamp demonstrates, we must examine details very carefully, since through them we can discover the history that is the crucial difference between areal and typological linguistics.

Another problem in a typological approach to the Balkan sprachbund in which Macedonian dialectology has an important role to play is the fact that a given phenomenon is grammaticalized to different degrees in different dialects. Here, too, an check-list approach can result in misleading conclusions. Thus, for example, while object reduplication per se occurs in all the Balkan languages, it is not grammaticalized, i.e. obligatory, to the same extent in different languages and even in different dialects of the same language. Moreover, Balkan Slavic territory shows the widest range of variation. The grammaticalization of object doubling is weakest in Bulgarian insofar as it is never obligatory except in expressions of negated existence, e.g. răkavica e tuk, ama răkatja ja njama ‘the sleeve is here but the hand is not’. The situation is similar in Greek, where the only construction in which object reduplication is obligatory is in phrases such as, ola ta kserεi ‘he knows everything’. In Daco-Romanian, the position of the object as well as the opposition human/non-human affect the obligation for object reduplication, while in Albanian a combination of position and topicalization determine object reduplication, e.g. Madje Papa e vizitoi Tiranɛn ‘Even the Pope visited Tirana’ but Papa vizitoi madje Tiranɛn ‘The Pope visited even Tirana’ (cf. KALULI 1999). In Macedonian, the standard language reflects the western dialects on which it is based insofar as reduplication is generally obligatory for definite direct and all indirect objects, but in the eastern dialects, as in adjacent Bulgarian, and Serbian
dialekt (and also in the Macedonian dialects of Gora in southwesternmost Kosovo [VIDOEŠKI 1986, MLADENOVIC 2001]), reduplication is conditioned by topicalization and is not as integrated into the grammatical system. Likewise, in Aromanian, those dialects spoken in southwestern Macedonia have the same rules as in the local Macedonian dialects, while the dialects in northwestern Greece follow Greek rules. Megleno-Romanian, on the other hand, has the same degree of grammaticalization of object reduplication as in western Macedonian, as opposed to the lesser degree of grammaticalization in the eastern Macedonian dialects with which it is in contact. In the Romani dialects of the Balkans, which in other morpho-syntactic respects are highly Balkanized, object reduplication is also pragmatically conditioned (see FRIEDMAN 2001). These data taken together suggest a Balkan Romance impetus and a southwest Macedonian nucleus for the phenomenon.

Likewise the dialectal facts of the classic feature of the formation of the future with a reduced form of ‘will’ are quite complex. Thus, for example, in studies where Geg and Tosk Albanian are differentiated, Geg is generally presented as using a future with ‘have’ (plus infinitive), e.g. kam me shkue vs Tosk do të shkoj ‘I will go’. The representation of Geg is only partially true, however. In the case of describing future formation in Geg, we can see mostly clearly the pitfalls of overgeneralizing morphosyntactic features in describing a dialect group whose defining features are phonological developments that took place long before the morphosyntactic developments in question. In Southern Geg, the Central Geg of Upper Reka, the Northwest Geg dialects of Shkrel, on the west bank of the Buna near Shkodër, as well as in the central North Geg of Puka, the classic ‘will’ future using invariant do plus subjunctive is the normal future tense. In Has and Luma, near Upper Reka, as well as in the North Geg of Mirdita, to the south of Puka, the ‘will’ future is in competition with ‘have’ and dominates it. In Kelmend, Plav, and Gusii, to the northeast of Shkodër, there is a more archaic ‘will’ future of the type also found in Romanian and the non-Torlak Što- and Čakavian dialects of BCS, i.e. with a conjugated form of ‘will’ (see FRIEDMAN 2005a and references therein).

Turning to Balkan Slavic proper, we see that in connection with the grammaticalization of the future-marking particle, Bulgarian is less consistent than Macedonian in a number of respects. In the Balkan conditional mentioned above, Bulgarian conjugates the ‘will’ auxiliary in the imperfect, whereas in Macedonian, the future marker is normally invariant. Nonetheless the northern Macedonian dialects and that of Galičnik are more conservative and keep imperfect marking on the ‘will’. We can also note here that in Romani dialects, the isogloss marking the Balkan type of future formation (particle from etymological ‘want’ + finite form) coincides almost exactly with the boundaries of the linguistic Balkans (BORETZKY & IGŁA 2004), which can otherwise be defined in terms of the Torlak isoglosses for

---

3 Even in the western dialects of Macedonian, the rules are more complex than indicated here, but the basic principle of higher degree of grammaticalization still holds.
Slavic, the extent of Albanian speech in the middle of the nineteenth century, and
the extent of Balkan Romance speech to the northeast of that.

Deictic pronouns do not figure much in Balkan linguistics, and with reason: deictic systems in the Balkans generally show resistance to contact induced
changes. Thus almost all Romani dialects both within and beyond the Balkans
preserve its a unique quadripartite system based on two binary oppositions—<i>a/o</i>
‘proximal/distal’ and <i>d/k</i> ‘general/specific’, e.g. Arli <i>adava, odova, akava, okova</i>. Albanian and Balkan Romance have two-term systems that are relatively old, although we know that Greek, Macedonian, and BCS (as well as western Bulgarian) reflect the older Indo-European three-term system found also in Latin, Sanskrit (and, we might had, Turkish). Eastern Bulgarian, however, except in the Rhodopes and Thrace, has gone in the direction of East Slavic by losing the marker of proximity. In this context, the Macedonian dialects of Lower Gora are particularly interesting, since they have a two-term system under Albanian influence and have lost the distal marker. Eastern Macedonian also has a two-way deictic system, like eastern Bulgarian, but, as in Lower Gora, it is the distal that is lost. While the Albanian influence in Lower Gora is clear, the situation in eastern Macedonia may be related to earlier contact with Balkan Romance. The matter is
in need of further investigation.

As mentioned above, according to <i>Van der Auwera</i> (1998), Bulgarian is the
most Balkanized of the Balkan languages (has the ‘highest score’ in his terms),
while Macedonian, together with Tosk Albanian, are according to the methodology/metrics employed, somewhat less Balkanized because (standard) Macedonian lacks stressed schwa while Tosk has a so-called infinitive construction of the type <i>për të</i> ‘for to’ plus participle. ⁴ Next in descending order are Romanian, Geg Albanian, Greek, then BCS and Romani (tied), then Turkish. This is displayed graphically in a map that represents the positions of the languages in an approximate geographic fashion, but <i>Van der Auwera</i> (262) states:

Note finally that the isopleth map is not intended to say anything specific
about the cause or diachrony of the convergence; Though Map 2 depicts a
core and a periphery there is no suggestion that the Balkan type started in the
core nor that all isoglosses necessarily include the more central languages.
For Map 2 Bulgarian does happen to be included in all isoglosses, but
Macedonian is only half-in for the [schwa] isogloss whereas Tosk Albanian,
which has the same score as Macedonian, is half-in for the [infinitive loss]
isogloss.

And just before that he states:

The agreement is not perfect, the reason being that other linguists
implicitly or explicitly base their judgment on slightly different feature lists.
Thus we see <i>Hamp</i> (1977: 281) saying that Macedonian is the most Balkan

---

⁴ Although <i>të</i> by itself is a subjunctive marker with various possible translations, the collocation <i>për të</i> generally corresponds to English ‘for to’. The literal gloss is now archaic or dialectal in English.
language, and in the feature counting of Campbell et al. (1986: 561) Romanian is the winner.

Having in mind that the Balkan sprachbund is precisely an historical phenomenon (HAMP 1977), however, then describing a language in the Balkans in terms of being more or less Balkanized should not be divorced from the process of (linguistic) Balkanization, which is a phenomenon of contact-induced linguistic convergence that has historic causes and is thus diachronically motivated. And it is precisely in the dialectological (and historical) facts that these processes can be traced and accurate representations achieved. In this regard, the choice and representation of features is not arbitrary, and their historical origins need to be understood.

In this regard, there are two serious problems with Campbell, Kaufman, and Smith Stark’s (1986:561) claim that Romanian is the most Balkan language and that Macedonian “lacks several of the areal traits.” The first is that they apparently did not know the Macedonian data well enough to realize that it has a grammaticalized perfect in ‘have’ (Standard Bulgarian does not), that it uses the same prepositions and adverbs for both location and direction, and that it has genitive/dative syncretism. The second is that disparate vocalic alternations in Romanian, Bulgarian, and Albanian (not Greek, pace the initial sentence in their presentation of the feature), which they call ‘Vowel harmony (or umlaut)’ involving the diphthongization of mid-vowels in Romanian, the ancient umlaut of /u/ and /a/ in Albanian, and the backing of /æ/ (later /e/) to /a/ in certain stressed syllables in [parts of eastern] Bulgarian are claimed as a Balkanism (1986:569). Crucial in the Bulgarian phenomenon is the fact that the original vowel is not fronted, rather, in the relevant dialects, a front vowel is backed except when followed by an historically palatalized consonant (itself the result of the front vowel in the following syllable). They also fail to take into account differences in relative degree of grammaticalization of, e.g., pronominal object reduplication. When the spurious Balkanism is dropped and Macedonian’s features are included (and here we include the majority of Macedonian dialects), then Macedonian, even using such a partial list, is indeed the most Balkanized.

A problem with typological approaches to the Balkan linguistic league that do not take into account diachrony was first voiced by Hamp (1977:279, 281) and cited in Campbell, Kaufman, and Smith Stark (1986:534-35):

Yet while the comparative method is unquestionably an historical study, the field of areal linguistics is no less so; for it too is occupied with analyzing the result of specific, if multiple, linguistic events of the past. Both the comparative method and areal linguistics are historical disciplines—twin faces of diachronic linguistics, if you will.

Hamp (1977:282) also observes:

Pavle Ivić has pointed incisively to the difficulty in drawing compact borders to a Sprachbund; the configuration is much more that of a spectrum. Yet here we have a multiple offender in Albanian, which in other ways seems to lie near the heart of the Balkan Sprachbund. A gross inventorizing would never
catch this important textural aspect. [...] [A]real questions can be approached meaningfully and fruitfully only if they are treated in specific terms for what they are—the results of developments with historical depth and specificity.

As can be seen from the foregoing, the Macedonian dialectal picture is considerably more complex than numerological or Eurological approaches to Balkan linguistics would suggest. In particular, in western Macedonia, where Balkan mutual linguistic interaction is most vigorous and complex, contact-induced change moves in different directions in and among different languages depending on location, feature, and sociolinguistic relations. As HAMP (1977) has observed, in the nineteenth century, when the discovery of the regularity of sound change was in progress, we did not have a precise understanding of the place of changes conditioned by contact phenomena vis-à-vis divergences within given groups. Now, however, we understand that divergence in a linguistic family and convergence in a linguistic league are nonetheless part of the same process and should be examined with the same rigor. We must therefore be careful to distinguish areal and typological approaches in general, and in Balkan linguistics in particular. This is not to say that a typological approach to the Balkan languages cannot afford insights into how language works. For example, as I have argued elsewhere (FRIEDMAN 2005b), the behavior of Balkan admirativity shows correlations between tense and discourse factors that may well be typological in nature. Once the genetic and areal facts have been carefully sifted, there is as much for typology to say about human language in the Balkans as elsewhere.

References
MIKLOSI, FRANZ (1861): Die slavischen Elemente im Rumanischen, in: Denkschriften der Kaiserlichen Akademie der Wissenschaften, Philosophisch-historische Klasse 12,1–70.

VICTOR FRIEDMAN
Department of Slavic Languages and Literatures
University of Chicago
Chicago, IL 60637 USA
vfriedm@uchicago.edu