The field of Balkan linguistics, like all areas of human scholarly endeavor, has developed in two contexts: the external context that defines it as a discipline of study (cf. Gal & Irvine 1995) and the internal context, which is to say the on-going conversation among its practitioners. The end of a century and, indeed, of a millennium, is certainly an appropriate time to reflect on past achievements and new directions.

In discussing the origins of Balkan linguistics as a discipline, we are actually faced with a number of choices. The 1770 trilingual dictionary of Theodore Kavaliotis and Thunmann’s 1774 history of East European peoples, which reproduced Kavaliotis’ dictionary and first suggested a link between the modern Albanians and the Illyrians of classical times, are taken by some as the starting point (Desnickaja 1970: 46). These works are certainly notable achievements in the birth of modern Balkan studies, as was the almost contemporary quadrilingual dictionary of Daniil of Moschoplis (Voskopojë), which was the first modern work that provides us with parallels of all four of the classic Balkan linguistic groups: Slavic, Romance, Albanian, and Hellenic. I should note that I am quite deliberate in my use of terminology that eschews the specificity of modern language names. At the time Daniil’s Tetrágllasson was first published in the late eighteenth century (the exact date is a bit problematic, see Kristophson 1974), none of the modern Balkan nation states (and official standard languages) pertaining thereto had yet come into existence. Daniil’s own terminology (Rðmikà, Vláhika, Alvanitika, Voulgárika) reflects an eighteenth century linguistic consciousness to which it would be anachronistic to apply late twentieth century conceptions.

Important though these lexicons are for our understanding of the state of Balkan dialects at this crucial time, however, from the point of view of theorizing the development of Balkan linguistics as well as historicizing the
development of linguistic theory, they represent raw materials. In this sense, it could be argued that Hesychius was the precursor to Kavaliotis, since he recorded a lexical specimen of Illyrian. I do not wish to suggest, of course, that Hesychius was the first Balkanist, nor do I wish to belittle the achievements of Kavaliotis and Daniil, but from the point of view of Balkan linguistics as an academic discipline defined by both geographic area and theoretical discourse, the works of Kavaliotis and Daniil remain sources of data but not theory. We could actually go one step further and note that given that the intent of these first comparative lexicons was the Hellenization of non-Greek-speaking Balkan Christians – see Tsitsipis 1998 on the ideological underpinnings of such efforts – they were quite in contrast to the comparative works of later periods (e.g., Miklosich 1861, whose interest was purely academic, or Pulevski 1875, whose purpose was just the opposite of Daniil’s).

I would argue that Kopitar’s 1829 formulation that Balkan Slavic, Balkan Romance, and Albanian give the impression of what I have translated into English as ‘one grammar with the three lexicons’ (Friedman 1997) – although still not an explicit theory of Balkan linguistics and omitting as it does Hellenic and based as it is on a relative paucity of data – is nonetheless the terminus ad quem of our field, insofar as it is the first statement to point explicitly to the grammatical commonalities that, as such, are key to the concept of areal linguistics – as opposed to typological or genetic linguistics (cf. Hamp 1977).

The next moment to be discussed is the formal recognition of the areal relationship of the linguistic league as analogous to the genetic relationship of the linguistic family. Schleicher’s 1850 formulation, which, using the ideological models of his day, refers to the Balkan languages as having ‘corrupted’ one another despite belonging to different families (i.e., groups), is taken by some (e.g., Simpson 1994: 210) to be the first such formulation. And as Sandfield (1930: 12) pointed out, it is Miklosich’s 1861 study of Balkan Slavic and Balkan Romance that contains the first explicit list of Balkanisms, i.e., some of the grammatical traits common to the four classic Balkan language groups (types of future formation, uses of subjunctive clauses, etc.). Nonetheless, it was Trubetskoj in 1923 in a Russian-language periodical and again in 1928 at the First International Congress of Linguists, who explicitly formulated the linguistic league as a theoretical concept.

It was during this same period that Sandfield published his classic study, first in Danish in 1926 and then in French in 1930, summarizing and systematizing the results of the considerable research that had been conducted until that time. In keeping with the practice of the period, almost half the work is spent on loanwords, and the remainder is defined as ‘outside the lexicon’. I should note that while lexical Balkanisms are certainly entitled to citizenship in the realm of Balkan linguistics, and most assuredly provide valuable – indeed sometimes unique – testimony to the history of linguistic contact, it is the grammatical rather than the lexical commonalities that constitute the definitive characteristics of the linguistic league as we now understand the concept.

Although the three or four decades that followed the publication of Sandfield’s summation were by no means fallow – many scholars, too numerous to mention here, were publishing specifically Balkanological studies during this period – including the short-lived Revue internationale des études balkaniques (1934–1938) as well as many significant works of the 1950’s (including Golab’s 1956 important observation of the isoglosses) – nonetheless, it was in the 1960’s that Balkan linguistic studies as such experienced a qualitatively significant upsurge that has continued unabated. One of the most important manifestations of the early years of that increased activity, and a major result of the research from the years preceding it, was volume VI of the proceedings of the first congress of AIESEE (Gáláb, Georgiev & Zaimov 1968). This volume stands as a kind of monument to the state of Balkan linguistics of that period, with virtually every scholar in the field at the time represented.

During the last third of this century, a rich trove of scholarship on Balkan linguistics in the form of focused books and articles has been produced, e.g., Joseph (1983) on the loss of the infinitives, Civić on syntax (1979), and Sawicka (1997) on phonology, to take three samples of three different types. A complete bibliography of just the 25 years since the period covered in Schaller’s (1977) bibliography would produce a volume many times the size of that earlier work. Numerous handbooks and surveys have also appeared such as those by Feuillet (1986), Schaller (1975), Solta (1980), Banfi (1985), Asenova (1989), and Demiraj (1994), as well as Desničkaja & Tolstoja’s (1990, 1998) edited language-by-language collection and Reiter’s (1994) eurolinguistically oriented work. We can add here Steinke & Vraciu (1999), whose manuscript, however, dates mostly from 1974. As the century closes, the long called-for Balkan linguistic atlas (e.g.,
Deanović (1961) is finally beginning albeit on a small scale with the University of St. Petersburg’s Malyj dialektologicheskij atlas balkanskix jazykov (Domoseckaja & al. 1997, Sobolev & al. 1997, 1998, 1999, Plotnikova 1996). It is worthy of note that while in the past Balkan linguistics have been discussed at general linguistic and phonetic conferences (e.g., Trubetzkoy 1928, Belić 1936), in recent years it is the International Congress of Slavists that has served most as the ‘home-away-from-home’ in terms of Balkan initiatives, as there is now a special standing committee on Balkan linguistics founded at the eleventh Congress, in Bratislava in 1993 (but cf. also the dictionary already proposed in Batowski 1939).

We must now say something about the external history of our discipline. It was precisely at the end of the eighteenth and beginning of the nineteenth centuries, when Western Europe first began taking notice of the Balkans and their languages in modern times, that speakers of those languages were first beginning the endeavors that would lead to the end of the political and social conditions that had created the Balkan linguistic league as such in the first place, i.e. the creation of Balkan nation states and their standardized national languages. One of the problems of much of modern Balkan linguistics, especially during this most recent upsurge in activity, is that it relies on data from these standards. It is thus using dialectological methodology on standardized data. Such an approach is not without merit, insofar as ‘Balkanization’ has penetrated so deeply into the total structure of all the relevant diasystems that even a standardized one will reflect this (e.g., Nikolaeva 1996; Aronson 1981, 1994; Kramer 1988, 1994), but in many instances the standards have purposefully, sometimes for ideological reasons, excluded specifically Balkan features as low style, sub-standard, or dialectal. The lexical example of Turkisms is the most obvious case in point (Kazazis 1972), but a variety of other features can be included here as well. Thus, for example, object reduplication, which is a typical Balkan syntactic feature, is subject to greater constraints in literary Bulgarian than in colloquial Bulgarian (Friedman 1994a), the leveling of relative pronouns – another Balkan syntactic feature – is greater in Macedonian dialects than in the literary language (Topolińska 1995). Studies such as Lindstedt (1998), which attempt to ‘weight’ or quantify Balkanisms, come out skewed in part because they do not take dialectal data into account; e.g., analytic accusative marking is typical of Balkan Romance (direct object marker pe) but, while absent from the standard Balkan Slavic languages, is also found in the west Macedonian dialects in closest contact with Aromanian. Aromanian itself is often missing from more superficial Balkan linguistic accounts, which are satisfied with Romanian as the single representative (and the Moldavian dialects almost never figure in any accounts). And within Aromanian, the dialects of the extreme southwest such as Beala di supră are considerably more balkanized (e.g. in possessing distinct evidentials and more simplified nominals) than those of other areas.

Topolińska (1998) makes the point that the Balkan Sprachbund as such no longer exists except as an historical artifact – rather like a given stage in the development of a genetic family. While it is indeed true that the creation of national boundaries has broken up the larger unity that constituted the Balkan Sprachbund as such, and moreover the standardization of national languages and the concomitant effect of mass literacy and education has had a similar effect on linguistic development, nevertheless, the same type of multilingualism with the same Balkan languages continues to exist at the communal level in all the Balkan countries. Despite increasing ethnic isolation (e.g., Icvecska and Salihu 1998) and aggressive assimilatory policies in various Balkan countries (various Human Rights Watch/Helsinki reports attest to this), which work against neighbor knowing the language of neighbor, there is still much that can be done to investigate both the remnants and the continuations of the Balkan Sprachbund.

Most works after Sandfeld have not changed the basic data and methodology but rather have at most expanded upon them. Of post-Sandfeld handbooks, Asenova (1989) goes the furthest in both expansion of the data base and development of methodology, but much of the dialectological collection of the turn of the century remains to be exploited. In this context we can cite as an example the Turkish part of Pulevski’s dictionary (1875), which, despite Hazal’s (1963) observation that it is a precious document of colloquial Balkan Turkish, remains to be exploited. (Studies of the Albanian part, e.g., Çable 1971, Friedman 1990, Jašar-Nasteva 1984, also could be expanded.)

In the context of our mention of Turkish, we can also raise the issue of those languages and dialects not usually included in Balkan studies. Thus, for example, Sandfeld (1930: 3) mentions Romani and Judezmo in a footnote, he does not use any data from those languages, although Miklosich’s 1872–1880 studies were by then available. Turkish is usually mentioned only as a lexifier and, occasionally, as a source of the so-called evidential. Although recently Balkan Turkish and Romani are beginning to receive more attention (e.g., Boretzky 1995; Igra 1996; Kostov 1973; Friedman...
1985, Forthcoming-a; Matras 1994, Uhlik 1973), there is still much to be done. Judezmo is still very much in the margins of Balkan linguistics, and in fact Joseph (1983: 252–253) demonstrates that in terms of one of the classic features, viz. infinitive replacement, Judezmo has remained relatively conservative. This may have to do with the relative lateness of Judezmo’s entrance into the Balkans, since large numbers of Judezmo speakers did not arrive in Ottoman Turkey until after their expulsion from Spain in 1492, i.e. well after both Turks and Roms had arrived. Moreover, Judezmo was in competition with Yavanic for at least part of the time after its arrival. There are also significant social factors that could account for the less Balkanization in Romani and Judezmo.

Unlike the other Balkan languages, which were the objects of two-way multilingualism, Romani and Judezmo were subjected rather to one-way multilingualism. In other words, speakers of the ‘classic’ Balkan languages (and also Balkan Turkish) learned other languages and heard their languages spoken by others. In the case of Romani and Judezmo, however, their positions as languages of stigmatized groups (see Friedman & Dankoff 1991), both of them strictly endogamous, meant that their speakers were of necessity multilingual but their languages were rarely learned by others. At the level of the mahala [neighborhood], of course, children did learn each others’ languages, so there were non-Roms who knew Romani and non-jews who knew Judezmo, but both such phenomena were relatively rare vis-à-vis the level of multilingualism among the other Balkan languages. The occurrence of Judezmo words and phrases in folk songs and folk poetry or of Romani words in slang or secret languages (e.g., Cvetkovski 1988: 190; Jašar-Nasteva 1987) does not contradict this principle but rather is an indication of the relative rarity of unidirectional multilingualism (as opposed to unidirectional multilingualism) in these languages. Cf. also codeswitching phenomena in Macedonian folk tales, in which Jews speak Turkish rather than Judezmo (Friedman 1995). Of these three languages (Romani, Judezmo, and Balkan Turkish), Romani is the most viable Balkan in terms of surviving multilingualism. The Jewish communities of the Balkans were mostly destroyed by the fascists during World War Two, and those that survived emigrated. Balkan Turkish is steadily losing ground both to migration and to other assimilatory factors (e.g. the dominance of Albanian among Muslims in Macedonia, cf. MILS 98.01.21, 98.02.13, MIC 98.02.03, RFE 2.15.2/98.01.23).

We should also mention here the issue of divergence versus convergence. In Friedman (1983), using Macedonian and Albanian data, I noted the fact that superficial resemblances sometimes mask underlying differences of structure. In Friedman (1978) I likewise note that superficial resemblances between Bulgarian and Turkish could be explained by contact-induced convergence rather than simple borrowing. Fielder (1999) has expanded on this. The different types of multilingualism mentioned above also involve differences between contact phenomena and boundary maintenance. This can be seen in the Macedonian dialects of Balkan Romani, where phonological conservatism seems to serve as a marker of ethnolinguistic boundary in the face of syntactic convergence (Friedman 2000), and within syntax the noun phrase seems more resistant to contact phenomena than the verb phrase (Friedman Forthcoming-b).

Given the interesting phenomena found in specific dialects of Balkan languages, e.g. the admixture of the Aromanian of Beala di suprâ (Friedman 1994b), but especially in view of the radically balkanized structure of western Rumelian Turkish (Friedman Forthcoming-a), it is unfortunate that we have no published, detailed studies of Balkan Circassian (Kânchev 1900: 116, 178, 215). They might have given some unusual insights one way or another. The Balkanization (or Indo-Europeanization) of western Rumelian Turkish is sometimes explained by accounts that the speakers are descended from Indo-European speaking converts to Islam rather than Turkish immigrants. The relative closeness of the Yürük dialects of Eastern Macedonia to eastern Rumelian Turkish (as well as these speaker’s own oral traditions – Jašar-Nasteva 1986) suggests that this might be the case. The relative grammatical conservatism of Judezmo is another case in point. It may be a matter of time depth (pre-vs post-Ottoman, cf. Romani, which is pre-Ottoman) but in any case, it is too late to test Circassian. Almost all the remaining known Balkan speakers were in Kosovo and were evacuated to Circassia (RFE 99.05.22, RFE 3.100.1/99.05.24) during the crisis period of the 1990's.

In addition to broadening the linguistic base, distinguishing carefully between dialect and standard, and making greater use of both older and newer dialect studies, Balkan linguistic studies need to integrate findings in other areas of linguistics into their methodologies and results. Thus, for example, Masica's (1976) mapping methodology could be fruitfully employed in revising and reaffirming earlier insights. In this connection we can mention what we can call ‘the European question’ raised by Reiter.
Interestingly enough, the quest for ‘universal grammar’ generated and dominated in the relatively recent past by the Massachusetts Institute of Technology’s faculty and followers does not pose a particularly fruitful avenue of investigation for Balkan linguistics. Joseph (2000) questions how insights obtained from this quest, oriented as they are to levels below that of surface structure ‘are revealing beyond what might be found if one were to compare any arbitrary set of typologically related languages chosen on a basis other than geography [...]’. He goes on to observe that since ‘lexical borrowing, a quintessentially surface-oriented phenomenon so widespread in the Balkans, can shackle off into construction borrowing and thus syntax’, it becomes ‘problematic to view the similarities in terms of deep syntactic features such as parameter settings [...] if some significance for the Balkan Sprachbund is to be claimed for [these similarities]’. He concludes that such studies ‘are interesting from the perspective of the Linguistics of the Balkans/Comparative syntax of the Balkans, but not from the perspective of Balkan Linguistics’/comparative Balkan syntax’.

In other words, studies of the Balkan languages that explain their commonalities in terms of universal linguistic features tell us nothing about the specific contact phenomena that actually led to the observed commonalities.

To expand on a point made by Lindstedt (1999), I would also note that Balkanists need to talk more to creolist and vice versa. Mufwene’s (1996) point about language ecology (a term borrowed from Haugen [1972] and used recently by Cyxun [1998]) and the social rather than structural features that are responsible for creolistics as such (Mufwene 1997) is applicable to Balkan linguistics as well. Thomason & Kaufmann’s (1988: 95) explicit unwillingness to treat multilingual contact situations in detail, while understandable within the limits of what they were trying to achieve, nevertheless leaves the phenomenon unexamined and uninterpreted.

Thomason’s (1997) recent collection demonstrates the problems contact languages pose for the genetic model (cf. here Topolinska’s [1998] problems with the areal model). The hypothesis that the structures of pidgins and creoles are qualitatively different from those of other languages is clearly disproved by the many detailed case studies of non-Indo-European contact languages in this collection. Its insights can be applied with profit to the study of the Balkan languages, although dialects of these same languages provide challenges for some of Thomason’s generalizations, including the notion of core vocabulary (which is, of course, a widely held idea). While meetings like that of AIESEE, where those of us who know
these languages can talk to one another, are always stimulating, and while meetings of linguists in general provide broader fora, there is a middle ground that has yet to be exploited. Balkanists, Creolists, and specialists from other geographic areas of contact (South Asians, Mesoamericanists) could fruitfully meet as Contact linguists.

A broadening of disciplinary contacts will help in the formulation of a new study of the linguistic mechanisms of Balkan contact phenomena, expanding both the data base and the theoretical tools that have been deployed until now. As indicated above, the data base should seriously include Romani (Gypsy), Judezmo (Ladino), and Balkan Turkish (both the West Rumelian dialects and Gagauz). The expansion of theoretical tools should include the following: 1) a careful differentiation between the social and structural roles of different types of bilingualism and multilingualism (including code-switching phenomena and language ecology), 2) discourse analysis and pragmatics, 3) variation theory as applied to the question of divergence versus convergence and paths of grammaticalization, 4) theories of language ideology, the development of standard languages, distinctions of register, and national versus local language practices. By examining both the structural and the social mechanisms of multilingual contact-induced change in the Balkans we can reconceptualize the Balkan Sprachbund to produce new knowledge both about how language contact works in complex multilingual situations in general and about the Balkans in particular.

REFERENCES


——. 1993. How ‘Balkan’ are the Balkan Languages? Paper delivered at the Annual Convention, American Association of Teachers of Slavic and East European Languages, Toronto 28–30 December. [Unpublished manuscript.]


—. Forthcoming. a. West Rumelian Turkish in Macedonia and Adjacent Areas. In: The Turkish Language in Contact. Amsterdam.


MIC = Macedonian Information Center (Electronic News Service).


—. 1872–1880. Über die Mundarten und die Wanderungen der Zigeuner Europas I–XII. Vienna.

MILS = Macedonian Information Liaison Service (Electronic News Service).


