LANGUAGE,
BLACKS
and
GYPSIES

Languages Without a Written Tradition
and Their Role in Education

edited by
Thomas Acton
&
Morgan Dalphinis

© The Contributors and Whiting & Birch 2000

WHITING & BIRCH
MM
A number of common themes emerge in these papers, such as the pivotal role of intellectuals whose situation may be far removed by barriers of education, multilingual literacy, social class and politics from the majority of an oral language community. Ellingworth's warnings about Einzelgänger are as firm as those of Dalphinis about super-literates. Finally, one can also draw conclusions from these papers about the need for some kind of linguistic theory to be present in the school curriculum itself. The old school grammars of English or Katharevousa have been discredited both as unscientific and as elements in the maintenance of oppressive social orders, but their disappearance has left a gap as yet unfilled. Unless children learn some way of understanding how utterances are structured, of how in different circumstances we can, without even thinking about it (until we are made to think about it at school) distinguish native speech from that of, say, a foreigner failing to reproduce native speech, then exercises like translation between languages, or even a conscious control of one's own tongue become impossible. Devonish in particular makes it clear that a new linguistics has to root itself in different vernaculars, and in education, for a multilingual society to flourish.

The processes which led to and continue to affect the formations of the modern literary languages of the Balkan peninsula have their parallels elsewhere in the world; yet the Balkans constitute a unique 'living laboratory', due both to the great diversity of languages, and to the fact that these processes are well-documented, relatively recent, and on-going. The relationship of linguistics to the developments in these languages has undergone a number of vicissitudes. Map One shows the provinces, regions and states of particular relevance to this article.

ORIGINS OF MODERN LITERARY LANGUAGES

We consider here six language groups: Greek, Turkic, Romance, Albanian, Slavic and Indic. Each of these groups is represented by one or more literary languages ranging from well-established to nascent. Some of these languages have claims to older written traditions such as Ancient Greek and Old Church Slavonic. These traditions were interrupted because, as the International Commission of Enquiry into the causes of the Balkan Wars wrote:

... the Turkish conquest came, levelling all the nationalities and preserving them all alike in a condition of torpor, in a manner comparable to the action of a vast refrigator (d'Estournelles de Constant et al., 1914, p.29).

This break is crucial in understanding the relationship of Balkan literary languages to contemporary languages without written traditions. After the 'refrigeration', the thaw entailed by the disintegration of the Ottoman empire faced the peoples of the Balkans with the task of creating literary languages on the basis of vernaculars which had completely lost touch with their literary
variants, if such there had been in the first place, to such an extent that they were operating, in essence without a continuous written tradition. As a result the problems were and are precisely those faced by any language attempting to establish itself as a medium of written communication, with all the socio-cultural and political significance thus entailed.

In addition, the extra-linguistic factor of ethnicity must be taken into account. Much modern thinking makes mother-tongue the most important determinant of ethnicity. In the Ottoman empire, however, religion was of primary importance as the definer of millet (nationality). Under this system all adherents of the Orthodox Christian Church headed by the patriarch of Constantinople were 'Greeks' while all adherents of the state religion of the Turkish Empire, that is Islam, were 'Turks'. The Greek Orthodox church gradually took advantage of its privileged position in Constantinople to eliminate the autocephalous Bulgarian and Serbian Churches (in 1765 and 1767 respectively).

Thus Bulgarians, Serbs, and all other Greek Orthodox peoples in European Turkey were treated as ethnic 'Greeks' who happened to speak some other language. Likewise the Moslem Albanians, Bosnians, Pomaks, and 'Torbesh' were all 'Turks'. This identification of religion and ethnicity bears not only upon the developments of previous centuries, but also upon current trends.

The definition of the terms 'literary' and language are problematic. Although both are linguistic, their definitions are largely extra-linguistic, that is, mainly political. It has been said that a language is a dialect with an army of its own. Thus, for example, the Germanic speaking peoples of Scandinavia have no less than three mutually comprehensible languages, while people in Italy, China and the Arab world speak mutually incomprehensible dialects of the same language. In the Balkans, as elsewhere, the situation is complicated in that linguistic and associated ethnic claims are intimately connected with territorial and other political claims. For this paper, however, I will accept the definition of a 'literary language' as one which is codified and used in all spheres of a given political or ethnic unit's national life (Close 1974, p.31). I will also use the term 'literary' to cover the terms 'standard' and 'national' applied to languages; these are important distinctions, but do not affect the considerations of this paper. I will define the following Balkan languages as literary: Greek, Turkish, Romanian, Albanian, Macedonian, Bulgarian and Serbo-Croatian. In addition I will discuss Romani, Aromanian, Moldavian, Croatian and Bosnian in so far as these shed valuable light on processes of development.

**POLITICAL CONFLICTS AND LANGUAGE DEVELOPMENT**

**Greek**

The standardisation of Modern Greek has suffered considerably at the hands of linguists due to the diglossia which their disagreements encouraged. Greek diglossia goes back to before the beginning of the Christian era when teachers
The Balkan States: The Formation of Literary Languages

he switched the whole country, virtually overnight, to a Latin alphabet orthography which he had devised in consultation with linguists. The change of alphabet greatly facilitated the process of eliminating or naturalising Arabic and Persian vocabulary, not only by cutting off the link with those languages, but also by rendering phonological irregularities predictable from the Arabic orthography opaque (Perry 1982, p.16). In 1932 Atatürk founded the Turkish Language Society and held the first Turkish Language Conference. His intention was to create a literary Turkish based on the common spoken language and purged of that part of its Arabic and Persian component which had not entered common usage. The Turkish Language Society was charged with creating new vocabulary based on dialect words, borrowing from other Turkic languages, and, most important, new derivations based on roots already present in the language.

The extremist purism of the Turkish Language Society during the 1930s, however, gave rise to one of the most bizarre linguistic theories ever to emerge from a standardisation movement, the so-called Sun Language Theory (Günes Dil Teorisi). According to this theory all human language originated from man's utterance of the primal syllable A(3), and all words of all languages can be derived by a series of formulas from this primal syllable. In addition it was claimed that Turkish was the mother of all languages (Tankut 1936). While this theory rendered most Turkish theoretical linguistic work of the period invalid, it did have the positive effect of tempering radical purism; for if Turkish was the mother language, then Arabic and Persian words in it were ultimately of Turkish origin, and could thus be claimed as native Turkish. It is even suspected that Atatürk himself launched this theory for the purpose of controlling the radical purists (Heyd 1954, p.34). It is in the role of purism that Greek and Turkish present an instructive contrast. The purism of Greek is archaising and politically conservative, that of Turkish is innovating and politically liberal or radical. During the most recent period of military rule in Turkey, when public discussion of politics was forbidden, right and left wing newspapers began to editorialise on linguistic usage. The former, led by Tercuman, attacked the Turkish Language Society as too radical; the latter, led by Cumhuriyet defended it. Both sides did so in the name of Atatürk, the former claiming that the Turkish Language Society had gone too far, the latter claiming that it was carrying on his mission.

Romanian

The case of Romanian differs from both Greek and Turkish. The earliest document in Romanian is a letter from 1521. With the Turkish conquest, Greek gradually became the dominant language in Wallachia and Moldavia due to the millet system, because the majority of the population was Orthodox Christian. The situation in Transylvania was considerably more complicated. After the reformation Transylvania had four official religions: Catholicism, Unitarianism, Calvinism and Lutheranism. The ruling classes had two languages: Hungarian and German. The majority of the population consisted
of disenfranchised Romanian-speaking Eastern Orthodox serfs. As education was tied to class and religion, there were no Romanian schools. During the eighteenth century, the Habsburgs introduced the Uniate Church into Transylvania, in an attempt to Catholicise the Romanian speaking majority and thus promote centralisation and the integration of Transylvania into their empire. The result was that Romanians were exposed to education, learned that their language was descended from the Latin of the ‘noble Romans’ who had ruled a vast empire, and thus acquired a new sense of dignity before their German and Magyar masters. The resultant movement, called Latinism, served as the basis of Romanian nationalism and the development of the Romanian literary language (Verdery 1983, pp.84, 108-121). The movement spread also to Wallachia and Moldavia. The subsequent rise of literary activity was particularly concerned with lexicographic considerations and the relationship of Romanian to Latin and the living Romance languages that already had literary traditions, especially French and Italian. This was particularly true in the coining of neologisms and the devising of an orthography: French was more important for the former and Italian for the latter. The dialectal base of the literary language which emerged was that of Wallachia, particularly Bucharest. There were early attempts to create a literary Romanian which combined Aromanian and Daco-Romanian dialects, but these were abandoned as impractical (Close 1974, p.67). In 1859 Moldavia and Wallachia were for all practical purposes united into an independent country, and the Cyrillic alphabet which had been used for Romanian since the sixteenth century was officially replaced by a Latin orthography. Transylvania became part of Romania between 1918 and 1920.

Unlike Greek and Turkish, in which the divisions have been language-internal, the divisions affecting Romanian have been external, Moldavian on the one hand, and Aromanian (Macedo-Romanian) on the other. As a result of the annexation of Bessarabia by the Russian empire in 1812, the principality of Moldavia was divided in half. The half that remained dependent on Turkey formed part of independent Romania in 1859. Bessarabia became part of Romania in 1919 and was re-annexed by Russia in 1940. During the inter-war period Russian claims to Bessarabia were historicopolitical: the ‘people’ of Bessarabia wanted to live in a communist state. After World War II, however, the Russians could no longer use this argument, since Romania was also a socialist country, and so the concept of a separate Moldavian nationality with its own Moldavian literary language was pursued in earnest (King 1973, pp.100-105). As Dyer (1985) has shown, however, literary Moldavian is based on the same Wallachian dialects that serve as the basis of literary Romanian. The only significant differences between the two literary languages is that the former uses a Cyrillic orthography and has a larger component of East Slavic loanwords.

Aromanian, the Romance language of the scattered groups living in Albania, Greece, and Macedonia, had separated from Daco-Romanian by the tenth century. While it could not be integrated with literary Daco-Romanian for a variety of reasons (Close 1974, p.67), the fact that it does not constitute a

The Balkan States: The Formation of Literary Languages

separate literary Balkan language is worthy of comment in the context of the pluralist policies of Yugoslavia, particularly Macedonia. It would appear that religion is still an important defining ethnic factor, in so far as all the Muslim non-Slavic minorities of Macedonia have or are obtaining a large measure of cultural autonomy, whereas the Aromanians, who constitute the largest Christian non-Slavic minority (0.5% of the population) do not have such rights.

Albanian

Like Romanian, Albanian did not have an ancient literary tradition. The earliest document is a baptismal formula embedded in a Latin text from 1462. While Albania was under Ottoman rule, an Albanian literature did develop among the Albanians who had fled to Italy, the Arbëresh, but this did not have much effect on events in Albania. The Albanians were divided among three religions, Catholics (10%), Orthodox (20%) and Muslims (70%), (Byron 1979a, p.17) and the millet system worked especially to their disadvantage. The Orthodox were subject to Hellenisation, the Muslims were considered Turks and therefore denied linguistic rights even after various Christian peoples had begun to gain theirs, while the Catholics were few, and largely isolated in the mountains of the North, although there was a significant community in Scutari (Skendi 1980, pp.187-204). The millet system also put the Albanians in danger of being completely partitioned by their Greek and Slavic Christian neighbours as the Ottoman empire was disintegrating, and so in 1878 in the wake of the Congress of Berlin, a group of Albanians formed the League of Prizren to promote Albanian nationalism, defend Albanian territorial integrity; and promote the Albanian language (the one national characteristic that cut across religious boundaries). A key issue was the choice of an alphabet. Each of the possibilities had religious implications: Arabic implied Islam, Greek implied Orthodoxy, and Latin implied Catholicism. Cyrillic was also used, especially in Macedonia, but was never a serious national contender. In the case of the Latin alphabet the choice of symbols and digraphs was of interest to the European powers, because they were sponsoring schools and seeking to extend their influence in the area. Thus, for example, Austria-Hungary and Italy, which were both publishing textbooks and sponsoring schools, purposefully supported differing orthographies (Skendi 1980, pp.218-220). In 1908 the Young Turks came to power and the Albanians were briefly permitted linguistic freedom. That same year the Albanian Alphabet Congress of Bitola (Monastir) was a key step leading ultimately to the current Latin orthography. Fearing that an Albanian literary language might lead to an independent Albanian state, the Young Turks attempted to suppress the Latin Alphabet, but in vain. Independence was declared in 1912, and from then till 1945 the main focus of linguistic effort was the dialect question. Albanian has two very different dialect groups: Geg (North) and Tosk (South). In the early years of independence people wrote more or less in their own dialects and attempts at dialect integration
were threatened by a high degree of artificiality, as for example in Faik Konitza's suggestion that the Geg indefinite article nji and the Tosk indefinite article nje both be used in literary Albanian, the former for feminine nouns and the latter for masculine, although in fact no Albanian dialect ever makes a gender distinction in the indefinite article (Byron 1976, p.50). Under King Zog (1925-1939), efforts were made to base a literary standard on the Southern Geg dialect of Elbasan, but serious standardisation did not progress until after the second world war.

At this point it should be noted that the region of Kosovo, which is predominantly Albanian, was awarded to the Kingdom of Serbia in 1913, and subsequently became part of Yugoslavia. In Albania the number of Gegs only slightly exceeds that of Tosks, but Gegs are the only Albanian group in Kosovo, and the overwhelming majority of Macedonia's 23% Albanian minority. After 1945 the Tosk dialect of Korçë gradually emerged as the basis of the standard in Albania, while Albanians in Yugoslavia continued to pursue standard Geg. The Yugoslav government tried to encourage the idea that there were two separate nationalities and languages (Albanski in Albania versus Shiptarski in Yugoslavia) but this failed. In 1968 the Albanians of Kosovo officially accepted the Tosk-based literary standard of Albania, thus producing gjilha unifikuar, the 'unified language' (Byron 1979b).

Kosovo intellectuals, however, keep length and nasality in their speech, and efforts have been made to introduce certain Geg features into the standard, e.g. the reflexive pronoun i vetë which expresses a distinction lacking in most Tosk dialects, but the literary standard is still Tosk in the vast majority of its features.

South Slavic

The South Slavs have been considerably more fragmented than the other language groups considered so far. The territory of modern-day Yugoslavia was divided between Austria-Hungary and Turkey at the beginning of the nineteenth century. Autonomy, independence, and unification were achieved in bits and pieces during the next hundred years. During the nineteenth century the literary languages of the West South Slavs (Serbs, Croats Bosnians, and others) were characterised by increasing unity while those of the East South Slavs (Bulgarians and Macedonians) tended towards division.

The differences among the West South Slavs were considerable. Serbs are Orthodox and use Cyrillic; Croats are Catholic and use the Latin alphabet. Croatian dialects diverge from one another much more than Serbian dialects. Croatian tends to borrow from Latin and German and tends to create neologisms whereas Serbian is more likely to borrow from Greek and Turkish and accept loanwords from other Slavic languages, e.g. Croatian kolodvor, Serbian stanica, 'station'.

A major dialectal division cutting across these lines is the reflex of Common Slavic *E (the choice for example between snig, sneg, and snijej for snow). In 1850 Serbian and Croatian linguists and intellectuals signed the književni dogovor ('literary agreement') in Vienna, agreeing to adopt a Herzegovina-based 'ijekovian-shtokavian' standard developed by the Serb Vuk Karadzic as a common literary language. This was a compromise for both sides brought about by the political necessity of having a common language. Vuk's language, however, met with stiff resistance in Serbia, where it was not officially adopted until 1868. The Serbs had been using a form of Church Slavonic called Slaveno-serbian, which had an established literary tradition and orthography, albeit no direct connection with the spoken language. Vuk not only rejected the vocabulary and grammar of this language, advocating instead the use of the vernacular, but he also reformed its Cyrillic orthography, eliminating obsolete letters, and, following the principle of one letter per sound, introducing the grapheme <j> from the Latin alphabet. This last move outraged many Serbs because of its Roman Catholic implications. The majority of Serbs, however, never accepted the ijekovian pronunciation, and ironically enough, Vuk's ijekovian standard came to be identified as Croatian. In 1954 a new agreement was signed in Novi Sad in which the various Serbo-Croatian speaking peoples of Yugoslavia reaffirmed their commitment to a unified literary language. In March 1967, however, a group of Croatian intellectuals issued a manifesto proclaiming Croatian as a separate language, and problems continue to this day. In recent years the Serbo-Croatian speaking Muslims who constitute the plurality in Bosnia-Herzegovina, and who are recognised as a separate ethnic group, Muslimani (now Bosniacs), have also been agitating for a separate codified standard distinct from both Serbian and Croatian. Further conflict is created by the prestige dialects of the Serbian and Croatian capitals (Belgrade and Zagreb), which both differ from Vuk's standard (Magner 1981). Thus the centrifugal forces of the nineteenth century have been replaced by centrifugal ones in the twentieth (Naylor 1980).

The situation of the East South Slavs, the Bulgarians and the Macedonians, has been quite different. As a group they were threatened by helleinisation during the first half of the nineteenth century due to the millet system, but with the establishment of an independent Bulgarian church (the Exarchate), different tensions began to manifest themselves. Two centres of Slavic literacy had arisen during this period: one in south-western Macedonia and one in northeastern Bulgaria. The people of the southwest had a distinct regional identification as Macedonians, and with the uncompromising attitude of the users of the northeast Bulgarian standard, especially after the establishment of the Exarchate, came the development of a separate ethnic and linguistic consciousness among many Macedonians. In dialectal terms, there is a relatively thick bundle of isoglosses, coinciding roughly with the Serbian-Bulgarian political border, which fans out when it reaches Macedonia, so that the dialects of Macedonia are transitional between Serbian and Bulgarian. Serbia and Bulgaria also had conflicting territorial claims to Macedonia (as did Greece), and they tried to support these claims by linguistic arguments, as well as by force of arms as in the second Balkan war of 1912 (Belić 1919, p.250, Vaillant 1938, p.119). The official Bulgarian attitude to this day is that Macedonian is a 'regional variant' of Bulgarian (B.A.N. 1978).
A comparison is sometimes made between Macedonian and Moldavian (King 1973, pp.100-102). The claim is that just as the Russians have fostered a Moldavian language in order to justify the annexation of Bessarabia from Romania, so the Yugoslavs have created a Macedonian language to justify their territorial claims against Bulgaria. There are, however, considerable differences between the two situations. Macedonian linguistic separatism is attested in print since 1878, and the first definitive outline of the bases for a Macedonian literary language dates from 1903, while Macedonian was still a part of Turkey (Friedman 1975). Thus, while the official recognition of a separate Macedonian language and nationality may well be in Yugoslav interests, the fact remains that this language and ethnic identification arose among Macedonians themselves quite independently of Yugoslav interests, before such interests even existed. The same cannot be said of Moldovan, since Bessarabia had already been annexed to Russia in 1812.

After the second Balkan war, Macedonia was divided among Serbia (later Yugoslavia), Greece and Bulgaria. In Greece, Macedonians have been subjected to gradual but unrelenting hellenisation. In Bulgaria, only the Bulgarian language has been permitted in Macedonia, except during the period 1946-1948 (from the end of World War Two to the Tito-Stalin break), when Macedonian was recognised in Bulgaria as a minority language. In Yugoslavia Macedonian was treated as a Serbian dialect between the two world wars, but literature was published and plays were performed in it as 'folklore and dialect literature'. On August 2nd, 1944, in keeping with Tito's pluralist nationalist policy, Macedonian was officially recognised as a separate literary language in Yugoslavia. The basis was the speech of the West Central dialect region, and a generation of young linguists was established. There was a brief period when some people believed waiting until a team of Russian experts could be brought in, but this proposal was not accepted (Friedman 1985). The orthography follows the principles of Serbian rather than Bulgarian.

Macedonian presents one of the few examples where linguists participated in a more rational arrangement of dictionary entries. Following the classical tradition, the codifiers of the literary language at first listed verbs by the first singular (present). In literary Macedonian, however, a verb can have one of three stem-vowels (/u/, /e/, /i/) in all persons of the present except the first singular (and third plural), where this opposition is neutralised (only /a/ occurs). An American linguist, Horace Lunt, convinced the codifiers to use the third singular, which clearly shows the present stem, rather than the first singular as the standard citation form (Lunt 1951). This not only gave the dictionary entry more predictive power, but also helped spread the standard use of stem-vowels, an area of considerable dialectal variation. A major problem for literary Macedonian is the fact that Skopje, the capital and principal cultural centre, is outside the West Central area and is subject to considerable Serbo-Croatian influence. Nonetheless, an entire generation of speakers educated in the literary language has grown up, and they can and do use it consistently.

Literary Bulgarian, like Serbo-Croatian and Greek, had to face threats from archaizers who wanted Church Slavonic to become the official language of the emerging Bulgarian state. Having overcome these (although Bulgarian orthography was not completely modernised until 1944) as well as the hellenisers mentioned earlier, the creators of the literary Bulgarian were faced with the task of integrating a number of divergent dialects. On the whole, however, having experienced increasing centrifugal tendencies in the nineteenth century and a definitive split in the twentieth, the two East South Slav literary languages are currently characterised by tendencies to unify around their respective standards, as opposed to the increasing fragmentation of West South Slavic.

Romani

Although literary activity in Romani dates only from the early years of this century, a number of attempts have been made to use it in at least some of the functions of a literary language in literature and education, in the USSR, Scandinavia and elsewhere. The past decade or so has seen a significant upsurge in Romani nationalism, and concomitant with these activities there have been renewed attempts to create a Romani literary language (that is, a unified, standard language). Activities within the scope of this paper had one of their most important starting-points in southern parts of the former Yugoslavia, which has the largest settled Romani population. Of particular significance was the publication of a bilingual edition (Romani and Macedonian) of a Romani grammar in Skopje in 1980 (Jusuf and Kepeski 1980). This grammar is a significant signal of the efforts to create a Romani literary language for use in schools, and can be compared with certain Macedonian works by Gorgi Pulevski from the end of the nineteenth century (Friedman 1975). Both reflect the rising national consciousness of their respective peoples, both are polyglot as a reflection of the linguistic situations of their respective users, and both reflect a lack of consistency and standardisation characteristic of the pre-codified stage of a nascent literary language. In the case of Macedonian, Pulevski's work, like other manifestations of Macedonian nationalism at that time, was lost or suppressed for many years, although the results achieved in 1944 were consistent with the beginnings signalled by it. In the case of Romani, Jusuf and Kepeski's grammar may be able to serve as the starting point for a Romani literary standard, at least in Macedonia and adjacent parts of Serbia and Kosovo.

The chief problems facing all attempts at a Romani literary standard are the integration and selection among divergent dialects and the expansion of vocabulary. Jusuf and Kepeski draw on the three main dialects of Skopje: Arilja, Dzhambaz and Burguzhi. The Arilja dialect is the oldest in Skopje and serves as the principle one for the grammar, but the authors' main approach to dialect selection is to avoid it: they randomly use different dialect forms throughout the text, including some of the paradigms. For the expansion of vocabulary, Jusuf and Kepeski have followed the practice endorsed by the World Romani Congress of borrowing words from Hindi. Unfortunately, they have not adapted these borrowings to Romani phonology; for example, they
spell words with voiced aspirates, which are quite foreign to the Romani sound system, for example, bhavi, 'consciousness'. Another important issue is that of the choice of orthography. Jusuf and Kepeski chose the Latin alphabet with a Yugoslav-based spelling, although they also supply a conversion table for a Macedonian-based Cyrillic alphabet which they use when citing Romani forms in their Macedonian text. The bilingual linguistic practice which is already well-established in Yugoslavia made this choice, however, internationally aimed option easier.3 We should also mention here the suggestion of Cortiade (1984) to overcome certain dialect differences by means of the use of morphophonemic symbols in the spelling system, e.g. orthographic řome (řomeča) for phonetic (romea)/(romea)/(romea)/(romea), 'Rom', instrumental case, where the various realizations of what was originally intervocalic ř/ are readily predictable. In other environments, however, there is a considerable variation in the treatment of ř/, and so ultimately a choice will have to be made. Cortiade has argued that one of the more conservative Balkan dialects should serve as the basis of literary Romani, since various innovations are predictable in terms of the original base, but not vice versa.

When we look at the factors which have led to the selection of dialect bases for the various Balkan literary languages, we see a combination of political, cultural and numerical considerations. Thus, for example, in the case of literary Macedonian, the choice of the West Central dialect was motivated by three major factors: it was the most distinct from both Serbian and Bulgarian, it was the single largest relatively homogeneous dialect region, and it was the most readily comprehensible by the largest number of speakers from other regions. In the case of Serbian, Vuk Karadžić chose his native dialect. The victory of literary Tosk over literary Gg as the basis of Standard Albanian is due at least in part to the fact that the majority of the leaders of post-war Albania were Tosks (Byron 1979a). In Bulgaria and Greece, the dialect in the first area to become independent of Turkish domination ended up as the basis for the literary language, although other factors (shift of capital in Bulgaria, diglossia in Greece) have significantly complicated the picture. Similarly, in Turkey and Romania, it is the dialects of the capitals which have had the prestige to serve as the bases of the literary languages, although in the case of Romania, the original impetus came from outside the literary region. In this context, it can be said that the considerable Romani literary activity outside of the Balkans will also have to be taken into account, but the precise outcome of interaction and integration remains to be seen.

We see therefore that the use of linguistics in the formation of literary languages and the affirmation of nationalism has had a wide variety of effects. Historical linguistics has been used by archaiizers and purifiers, especially in the cases of Greek and Turkish. On the other hand dialectology has been the key to efforts of separatism and unity among the Slavic, Romance and Albanian languages. To these latter can now be added Romani. In the fixing of orthographies, phonology has played a role in all but Greek, which has retained its historical spelling. All of the other languages claim 'phonetic' spelling, although in fact phonemic would generally be a better term. Of the

The Balkan States: The Formation of Literary Languages

Slavic languages, Bulgarian is distinguished from Serbo-Croatian and Macedonian by a greater tendency towards morphophonemic spelling and as a result of the 1944 spelling reform, Bulgarian also has a few morphologically conditioned pronunciations of a type not found in Serbo-Croatian and Macedonian. Another example of a phonological phenomenon affected by standardisation is the automatic devoicing of underlying voiced consonants in final position. This is absent in Greek, Romanian and Serbo-Croatian, but uniformly present in Macedonian, Bulgarian and some dialects of Albanian and Romani. The approach in the Slavic languages is to portray the underlying phoneme orthographically, e.g. narod/narodi, 'people/people'. In Turkish, after considerable vacillation, the current orthographic practice favours phonetic representation, e.g. cep/cebm 'pocket/my pocket'. In the case of Albanian, the central dialects are characterised by final devoicing, a feature absent both further north and further south. In this instance language planners eventually opted for spelling final voiced consonants, e.g. zog/zogu, 'bird/the bird', despite the fact that this does not represent the pronunciation of the Tosk region whose dialects form the basis of the standard language. The spelling of final voiced consonants is helping to spread the pronunciation to areas where it is not native. On this point as on others Jusuf and Kepeski's Romani grammar uses alternatives without selecting between them e.g. in the spellings dad and dat, 'father'. The general tendency in most works written in Romani so far, however, has been to spell with voiced final consonants.

In the realm of morphology we have seen how the alternatives offered by different dialect forms can be used to enrich the literary language, or to create artificial distinctions. In the case of syntax and lexicon, the dialects (and earlier stages of the language) can be used to enrich the literary language; a key issue for these, however, (which does not affect phonology and morphology in the same way) is the extent of foreign lexical borrowing and syntactic imitation of foreign models. This has been an especially important issue for Macedonian, where the tremendous influence of Serbo-Croatian after 1945 threatened to alter its syntactic patterns significantly. Concerted efforts on the part of both codifiers and users of the literary language have succeeded in reversing this trend, however.

CONTEMPORARY SOURCES OF AUTHORITY IN LANGUAGE CODIFICATION

Each of the socialist countries, Albania, Bulgaria, Romania and each republic and autonomous region of the former Yugoslavia has an academy of sciences and an institute of language. The institutes are responsible not only for linguistic research but also for developing and defining literary norms and publishing grammars, dictionaries and other authoritative works. Each institute has its journal in which both theoretical and practical linguistic questions are addressed. The Institutes are not the only sources of normativisation, however. There are also Teachers' Unions and other language
organisations which publish linguistic or language journals on a more popular level and which are devoted to practical considerations of usage and codification. Similar linguistic topics are also discussed in the daily press, and there are many popular books on 'language culture', that is, normative usage. As a result, the codification of literary languages has developed in part through dialogue between codifiers and users.

In Greece and Turkey, language codification has been more closely tied to political parties, and linguistic tendencies are associated with political tendencies. (This is not to say that language was divorced from politics in the socialist Balkans, but only that codification has not been subject to the same types of vicissitudes in a one party system as in a multi-party system.) Although the Turkish Language Society functions in a manner similar to that of the institutes described above, neither Greece nor Turkey has quite the same type of institution due to differences in sociopolitical structure, and the Ministry of Education frequently plays a more active role in decision-making.

In future language planning and standardisation in the Balkans, linguistics can be used to advance compromise solutions to specific problems, choices or interpretations, or it can be used to justify divisive tendencies which are always potentially present. In Greek and Turkish, special problems are created by the strength of archaising and puristic tendencies, while the Slavic and Albanian literary languages face special difficulties due to the fact that their cultural and political capitals are located outside the region where the literary dialectal base is native. In Romani the chief problems which must be faced are dialectal compromise and vocabulary enrichment, and as these processes are still in their early stages, there is much that linguistics and the previous experiences of other languages can contribute.

NOTES

1. In this paper, I take the most widely accepted definition of the term 'Balkan', that is, the peninsula comprising the modern nations of Albania, Bulgaria, Greece, Romania, the former Yugoslavia and European Turkey (Turkish Thrace). A more detailed treatment of some of the themes of this paper appear in my article 'Linguistics, Nationalism, and Literary Languages: A Balkan Perspective' in Raskin V. and Bjorkman P (eds. 1986) The Real World Linguist: Linguistic Applications in the 1980s, Ablex, New Jersey, pp.287-305. A number of changes have occurred since this paper was originally written, especially in former Yugoslavia. Nonetheless the basic facts are accurate.

2. The former Soviet Union provides an instructive contrast. There, Romani, like all other national or literary languages except Estonian, Latvian, Lithuanian, Armenian, Georgian, and Yiddish, had to use a Russian-based Cyrillic alphabet.

REFERENCES

B.A.N. (1978) 'Edinstvo na bulgarskaja ezik v minalo i dnes', Bulgarski ezik, 28(1), pp.3-43
Byron J. (1979b) 'Language Planning in Albania and in Albanian-speaking Yugoslavia', Word, 30(1-2), pp.15-44
Friedman V.A. (1975) Macedonian Language and Nationalism in the Nineteenth and Early Twentieth Centuries, Balkanistica, 2, pp.83-98
Heyd U. (1954) Language Reform in Modern Turkey, Oriental Society, Jerusalem
Lunt H. (1951) 'Morfolojiayia na makedonskij gajgolj', Makedonski jazik, 2(6), pp.123-131
Tanku H.R. (1936) Güne-dil teorisi ne il dil tethiheri, State Publishing House, Istanbul