Romani in the Balkan Linguistic League

Victor A. Friedman
University of Chicago

1. Introduction

The Romani language raises serious and interesting theoretical questions for the concept of the linguistic league in general and for the concept of Balkan linguistic league in particular. In this paper I wish to examine some aspects of Romani as a Balkan language and the question of Balkanisms in general. The Balkanization of Romani must be understood on several levels. From a strictly structural point of view, we can distinguish three levels of Balkanization in Romani: (1) the earliest pre-European diaspora which resulted from contact with Greek (but also with Slavic), (2) the continued contact shared by all those dialects that remained in the Balkans, whether on Austro-Hungarian or Ottoman territory, (3) the intensive contact of sedentarized dialects. Moreover, a striking feature of Balkanization in Romani is that while it has obvious effects on certain aspects at all levels from phonology through syntax, there are other areas that in the canonical Balkan languages (i.e. Albanian, Greek, Balkan Romance, and Balkan Slavic) show the effects of Balkanization but in Romani do not. I would argue that this might not be for structural reasons but rather for sociolinguistic ones. Owing to the specific social position of Romani, resistance and boundary maintenance play a special communicative role that results in the differential adoption of Balkan features. Moreover — and this is a general problem in modern Balkan linguistics, especially regarding the Balkan Turkish dialects — the difference between more and less heavily Balkanized Balkan Romani dialects points to a general need for greater subtlety of analysis in the representation of Balkan language contact (cf. Joseph 1983).

2. Early contact

In discussing language contact, the source of a given phenomenon is always an important question: Is it the result of a borrowed pattern (calque), of interactive interference, of typological universal tendencies, or of a push in a direction in which the language was already heading? And how exactly is a linguistic league to be defined? The fact, for example, that object reduplication is characteristic of colloquial western Romance, raises questions concerning its relation to Balkanization in terms of areal restrictions. Masica (1976) has addressed this type of question for South Asia with interesting results. Given that Romani has been described — and not without justification — as a Balkanized Indic language (Matras 1994, 1995), let us take a moment to examine Masica’s discussion of South Asia, i.e. India sensu largo, as a linguistic area.

2.1 The Indic-Altaic Area

Masica (1976) investigates the question of the South Asian linguistic area.
by taking those features that are ascribed to it in various sources and tracing them as far as they will go on the Afro-Eurasian land mass and adjacent islands. Masica's methodology yields an interesting and perhaps unexpected result. When all the possible features are carried to their furthest geographical extent, what emerges in terms of a distinctively defined region is an Indo-Altaic hourglass (Masica 1976;180-81), i.e. a region comprising roughly the Kazakh steppe and the Indian peninsula connected via a narrow band traversing the Hindu Kush and the Pamirs. The features that end up defining this region are given in below and their presence or absence in Romani is indicated.

2.1.1 Second causatives

Although Romani has morphological causatives, double causatives are present only in dialects in contact with other languages that have the phenomenon, e.g. Hungarian. Such forms, however, use native material, e.g. darel 'be afraid' > daravel 'frighten' > daravel 'make someone frighten someone else' (Hübschmannová 1996).

2.1.2 Adjective-Noun order

This feature is found throughout the Balkans (except Albanian) and much of Europe (except the Italo-Celtic area): o baro kher 'the big house'.

2.1.3 Past gerunds (conjunctive participial devices)

This is absent from Romani, where only a present gerund occurs, e.g. phirindor 'while walking'.

2.1.4 Explicator-compound verbs

Romani usage is not the same as in South Asia. For example, the use of lel 'take' to mean 'begin' is a semantic equivalent, but the South Asian phenomenon involves a past gerund plus an auxiliary (e.g. Uzbek IP + ol-). The type of construction found in Romani occurs throughout the Balkan languages.

2.1.5 Dative-subject constructions (implies lack of 'have')

Romani has some oblique subject constructions (e.g., ladz si man 'I am ashamed'), but often the accusative rather than the dative is used. Most dialects express 'have' by means of an oblique construction, usually 'be' plus accusative or, more rarely, locative or dative, e.g. si man 'I have'. The use of tere 'hold', grasp' in parts of Rumelia is a later innovation, probably based on Greek. Despite the presence of lexical verbs meaning 'have' throughout the Balkans, most Balkan Romani dialects retain the older oblique construction.

2.1.6 Object-Verb word order (implies postpositions)

Verb final word order is permissible but marked. The normal order is Subject-Verb-Object. Me na bešav ki Skopja or Me ki Skopja na bešav 'I do not dwell in Skopje', O melalo pani na plena le ni o diungale ruva 'Even the wicked bears do not drink the dirty water', me čhavo te dikhav 'in order to see my child'.

Romani postpositions survive only as a small set of agglutinative case endings, otherwise Romani is basically prepositional (see Friedman 1991).

2.2 Roman and South Asia

Romani is in a special position, having migrated from the South Asian Sprachbund to the Balkan one. At the time the ancestors of today's Romani-speakers left India, their language had already undergone those influences shared by the early medieval Prakrits that had emerged up to that point. When we look at the modern Romani dialects in terms of South Asian areal features, we see that out of the six features one (adjective-noun order) is consistently preserved, one (past gerunds) is totally absent, and the remainder are present either as traces or remnants.

3. Romani and the Balkans

Turning now to Romani in its Balkan context, we can say that the influence of the Balkans is clearly greater in those dialects that remained there, and as Boretzky (1996) has shown, those that left the Balkans have to some extent lost, altered, or perhaps never developed some of those early, distinctively Balkan features. Thus, for example, although so-called infinitive loss is well known throughout Romani, some dialects outside the Balkans appear to be in the process of developing new infinitival constructions (see Boretzky 1995). Similarly, the definite article — a Balkan development shared by most Romani dialects — has virtually disappeared from certain Northern dialects in Russia, Poland, and the Baltic states (Ventcel' and Cerenkov 1976:329). The future using a reduced form of 'want', e.g. ka, is restricted to dialects of the Balkans and some Vlah dialects spoken outside the Balkans. There are also Vlah dialects spoken outside the Balkans that have lost this feature in contact with other Romani dialects (Hancock 1995:99-100). Thus it is useful to distinguish Romani as a Balkan language from Romani as a Balkanized language.

In discussing Romani in contact with the Balkan languages, we are faced with a specific paradox. On the one hand, Romani serves as a boundary marker distinguishing the Rom from the gadjo, on the other hand, the fact that all Rom will of necessity be bilingual by the time they reach adulthood and the intensity of the language contact cannot help but have its effects. In the context of the Balkans, what is striking is the difference between those features that have been assimilated and those that have not. I would suggest that certain features, (stress, intonation, parts of morphology and syntax) serve as ethnolectal markers (cf. Herson-Finn 1996) while others serve as connectors thus reflecting the dual purpose of language for both communication and differentiation. A study of these in Balkan Romani may throw light on similar phenomena in other Balkan languages.

3.1 Romani and Balkan Contact

As Rusakov (Eloeva & Rusakov 1990:8) points out, a difference between Romani and the other languages of the Balkans is the fact that the characteristics of the Balkan linguistic league developed in environments of mutual multilingualism, i.e. speakers of a given language were influenced not only by the structures of the other languages they learned but also by hearing their languages spoken as second languages. In the case of Romani, however, (like Judezmo), the nature of sociolinguistic relations was such that multilingualism went only in one direction. Although Romani-speakers learned other languages, their own language was not normally learned by others. This is a significantly different picture of interference than that represented, for example, by Macedonian-Arumanian bilingualism as posited by Golub (1984:9) for southwestern Macedonia, which is the area of the strongest and most consistent Balkan convergences. The following diagram of Arumanian-Macedonian calquing of the perfect (using 1 sg 'dine') illustrates this type of relationship (after Golub 1984:135):

Stage I. sum veezeral // anųcinață
Stage II. sum večeră/ imam večerano [borrowed] > amu cinată
Stage III. imam večerano/sum večeră < esku cinatu [borrowed]
/ amu cinată
Stage IV. imam večerano/sum večeră [re borrowed] < esku cinat/amu cinată

3.2 Romani and Balkanisms

The very nature of what constitutes a "Balkanism" continues to be the topic of lively debate, and Joseph (1983) has shown that even so classical a feature as Balkan infinitive loss is by no means monolithic within the Balkans. At the same time, however, he demonstrates that precisely Macedonian represents the most consistently Balkanized type in this respect. (Cf. also Friedman 1986, showing Romani as midway between Romanian and Albanian, on the one hand, and Slavic and Greek, on the other, with regard to this feature in typological terms). Nonetheless, starting with Sandfield (1930) and continuing up to Asenova (1989) and Demiraj (1994), there are features that are consistently identified as Balkanisms. It is these features that we shall consider here, with a focus specifically on Macedonian Romani as the Balkan dialects of the most Balkan region (cf. Hamp 1989, Gołąb 1984:6-10).

What follows here is a list of features based mostly on Asenova (1989) although both Sandfield's (1930) seminal work and Demiraj's (1994) more recent synthesis have also been helpful.2 The examples have been taken from Kostov (1973), Jusuf and Kepeski (1980), Jusuf (1996), Jusuf (1974), Elowa & Rusakov (1990), Matras (1994), and Uhlik (1973). Data from Sampson (1926/1968) and Holzinger (1995) were used to supply additional material. The plus indicates that the feature is a Balkanism in Romani or at least a feature shared by Romani and the canonical Balkan languages, while the minus indicates a Balkanism (or partial Balkanization) lacking in Romani. A sign in square brackets indicates either that it is not a complete Balkanization or that it is not completely present in Romani, i.e. it is limited either systemically or in terms of the dialects in which it occurs.

3.2.1 Keeping the vocative (+)

This is an Indic inheritance whose conservation may have been encouraged by Balkan contact. e.g. phral 'brother' phrala 'O brother'. Those dialects still located in the Balkans also borrow vocative forms of loanwords, e.g. in Arli tetka (from Macedonian) 'aunt', voc. tetko.

3.2.2 Double accusatives (+)

This is clearly a Balkan innovation, although in fact the Romani situation goes further than that in the canonical Balkan languages, e.g. ka dav tumaro 'I will give you bread', si man maro 'I have bread'.

3.2.3 Replacement of infinitive (+)

This is a classic Balkanism that is, however, gradually losing ground in some dialects spoken outside the Balkans, e.g. mangav te dīav 'l want to go'.

3.2.4 Universal relativizer is 'where' (kaj) (+)

Another Balkanism (or Hellenism) shared by all of Romani, e.g. me dīanav kaj ka den maro 'I know that you will give me bread'.

3.2.5 Analytic cases (+)

Although this tendency occurs in various Romani dialects, in the Balkans it is especially characteristic of Macedonian Arli, which has a longer history of sedentarization, e.g. ki jekh aindē versus jekha aindēate 'in a field'.

3.2.6 Analytic comparative and superlative (+)

This is a typical post-Byzantine Balkanism, i.e. one that developed after the first wave of Romani speakers left the Balkans for northern Europe. Those dialects that did not remain in the Balkans either maintain the old synthetic comparative in -eder with greater consistency, or devise other means to express comparison, e.g. baro 'big' pobaro, mābaro, dafha baro 'bigger', nābaro, embero, mābaro 'biggest'.

3.2.7 Future in 'will' (+)

Like the analytic comparative, this is another post-Byzantine Balkanism that has also been lost in some Vlax dialects that subsequently left the Balkans, e.g. me kā dīav 'I will go'.

3.2.8 Conditional = future-in-past (+)

This is another post-Hellenic Balkanism that survives only sporadically in the Vlax dialects, ka dīavā sine - ka dīavas - kamas dīav 'I would go'.

3.2.9 Turkisms (+)

Turkisms in Romani are for the most part limited to those dialects that remained in the Ottoman Empire, e.g. laf 'word'. They are almost entirely absent from the dialects that were not in the Balkans in the post-Byzantine period and occur in Vlax dialects only insofar as those dialects borrowed them via Romanian.

3.2.10 Future marker of necessity and negation is 'have' (+)

The following is a typical example: si man te dīav 'I have to go'. Some dialects also have a Balkan Slavic calque, e.g. nae te bitarav 'I shall not forget'.

3.2.11 Definite article (+)

The Roman definitive article is not postposed, as is also the case in Greek, e.g. o chavo 'the boy'. Uhlik (1951, 1973) mistakes morphologically marked borrowings in -va in dialects such as Gurbet for an article, e.g. beljava = belja 'trouble'. Although the suffix may be related to definiteness in Romanian, it has no such grammatical nuance in Romani. Sampson (1926/1968:152) argues convincingly that the Romani article is based on native pronominal material, given the preservation of oblique forms in l- < *t-. Thus, while not a borrowing from Greek, its placement is nonetheless like that of Greek rather than of the other Balkan languages.

3.2.12 Locative-dative merger (including ubi = quo) (+)

This appears to be limited to the non-Vlax Balkan dialects, e.g. me ka bešav/čav kiso Skopla 'I will live in/go to Skopje', kaj ka da 'where will you go', kaj beše 'where to do you live', t[e] = k[e] 'in, at, to'.

3.2.13 Object reduplication (+)

Although found in all Romani dialects, resumptive pronouns are limited to possessive and relative constructions and pronominal emphasis, e.g. O Rom kaj dikhjum ole 'The Rom that I saw [him]', Jek da sin la du 'he have 'A mother had two children', man si man jekh kher 'I have a house', man ma axmiz man kidisave bude 'don't mix me up in such affairs'. Romani object reduplication resembles Balkan object reduplication taken as an areal feature, which itself is not a uniform phenomenon (see Friedman 1994a).

3.2.14 Lexisymtoms with prepositions (+)

Romani participates only partially in such Balkanisms. Thus, for example,
Romani uses the dative in meanings of Slavic za, Albanian për, Greek για, Romance prentru, including the equivalents of English 'for' and 'about', and Romani uses te like Slavic da, Albanian të, Greek na, and Romance sə, but only some Balkan dialects have ki re, a collocation that is the equivalent of za da, për të, gia na, pentru sə. Similarly, Romani can use the ablative or its prepositional equivalent tar-like Slavic od - ot, Albanian nga (~ prej), Greek apo, Romance din (de, de la), e.g. tari xoli mejkal ime tara po vas 'from anger she drops the broom', hari potenari tari Dudjia 'a little younger than Dudjia', chinidu - mudardu tari butu 'dead tired from work', O Muto dolel e Ismeti tara ko 'Muto grabs Ismet by the ear', so posivo tari lađa nacela pe khereshe 'she runs home as quickly as possibly out of shame'.

3.2.15 **Vowel/prosody reduction and consonantal phenomena ([-])**

Length distinctions occur only in some Romani dialects spoken outside the Balkan area. However, these distinctions appear to be secondary rather than continuations of Middle Indic length. In any case, loss of prosodic length is characteristic of all Balkan Romani dialects, just as it is characteristic of the Balkan languages (although both Geg and southeastern Macedonian show the development of new length). An interesting feature of Romani prosody is the fact that stress placement remains quite conservative in the Balkans, whereas elsewhere there is a tendency to adapt to the stress of the major contact language. Thus most Balkan Romani dialects have oxytonic stress in native words, non-oxytonic in foreign words. This is especially striking in Macedonia, where the general tendency is to shift the stress away from the final syllable. To this we can add some local consonantal phenomena. In the Romani dialects of Macedonia as in both Macedonian and the dialects of other languages in Macedonia, e.g. Albanian and Turkish, we find final devoicing and the loss of /h/, e.g. dad - dat 'father', haljovel - aljovel 'understand' najhari - najari 'littlest', bulho - bulo - buvlo 'wide'. Similarly, the Romani dialects of Greece and Turkey reflect the Greek and Turkish universalization of dental and palatal affricates (only the dental /t/, /d/ in Greek, only palatal /ʃ/, /ʒ/ in Turkish, and also in the respective Romani dialects), whereas most of the Romani dialects of Macedonia and Albania differentiate strident and mellow palatal affricates as well as dental, at least at the phonetic level.

3.2.16 **Stressed schwa ([-])**

This is a 'partial Balkanism' since it is lacking in Greek, West Central Macedonian, and Geg. In Romani, too, schwa is marginal, occurring in loanwords or as a variant, e.g. gandinele 'think', sastrun/sastrin/sastron 'iron'.

3.2.17 **Teens as numeral-on-ten ([-])**

Here Romani, like Greek, does not follow the Balkan pattern of Slavic origin of the type ‘two-on-ten’, rather we have deš-u-daj ‘twelve’ (ten and two).

3.2.18 **Genitive-dative merger ([-])**

Romani is quite conservative in this regard: čhaveskoro (gen.) vs čhaveske (dat.) 'son'

3.2.19 **Dative as possessive ([-])**

Romani uses the genitive consistently.

3.2.20 **Aorist-perfect opposition ([-])**

This opposition is lacking in Romani, as it is on part of Romanian territory.

3.2.21 **Other**

In addition to the phenomena listed above, there are numerous examples of calques on Macedonian in Arli including some grammatical innovations that can be taken as Balkanisms, e.g. te najas Rifat e romske skalv = Macedonian da ne beše Rifat imeto na romot 'Was the Rom’s name by any chance Rifat?', vakerel korkori pea = Macedonian zaboruva sam so sebes 'he speaks to himself'. Cf. also kerava sibe 'I was doing', i.e. the formation of the Arli imperfect by means of the long present plus the invariant third person imperfect of 'be', which seems to recapitulate the original morphology of the imperfect (long present plus /s/) but also seems to calque the use of Macedonian third singular imperfect beše as a generalized emphatic past marker, e.g. beše sum bi 'I was'.

4. **Romani Balkanisms and Dialectology**

There are a number of other features that have been proposed as characteristic or partially characteristic of the Balkans, but those given here are sufficiently illustrative for our purposes. The Balkanisms of Romani can be viewed in terms of their relationship to Romani dialectology on the one hand and to Balkan linguistics on the other. In the first case we can distinguish those which are common to all of Romani as opposed to those found only in the Vlax and non-Vlax dialects of the Balkans versus those found only in some Balkan (especially non-Vlax) dialects. In the second case, Balkanisms can be classified as those which Romani shares with all the Balkan languages vs those which are only “partial Balkanisms” in Romani or the Balkans or both. Thus, for example, infinitive replacement is, ceteris paribus, common to all the Balkans and Romani and can be said to date from the earliest period of Greek contact. On the other hand, future formation with ‘will’ is a major Balkanism that only the Vlax and Non-Vlax Balkan dialects share, although we cannot know whether or not other dialects had the feature and subsequently lost it. The definite article is an example of a pan-Romani feature that is nonetheless only partially Balkan in the same way as in Greek: The article is present, but the position is not post-possed. The large numbers of Turksisms constitute a similar example, insofar as the Turkish lexicon is typical of all the colloquial Balkan languages — including the Balkan dialects of Romani — but disappears from Romani outside of the Balkans (cf. Friedman 1989). The uses of future-in-the-past as conditional and the negative existential/possessive as a negated future marker are typical Balkan calques in non-Vlax Romani Balkan. In a sense, Romani de-Balkanizes upon leaving the Balkans. This is certainly the case with regard to the infinitive, although the situation with the future is essentially moot. Interestingly enough, in the matter of prosody there is a tendency for the Balkan dialects to retain their distinctive (and non-Balkan) oxytonic stress, whereas outside the Balkan area there is a tendency to assimilate to the stress pattern of the contact language.

4.1 **Romani Reduplication**

Reduplication is a peculiar phenomenon within the context of the Balkan linguistic league and beyond it. It enhances communication and is especially characteristic of colloquial style. It is eschewed in literary languages except when they seek to be specifically colloquially based. Moreover, the phenomenon itself
varies from pragmatically conditioned to grammaticalized in different parts of the Balkans. The limitation of reduplication in Romani reflects the twofold role of communication and boundary marking. Matras (1996) on the basis of Romani data makes precisely the same claim as Friedman (1994a) on the basis of canonical Balkan languages — especially as illustrated by Albanian, Macedonian, and Vlah data — all of which is foreshadowed by Prince (1988), namely that pragmatic features can be borrowed in the course of language contact, and moreover they constitute the entry points from grammatical, i.e. structural change.

5. Conclusion

Differences in Romani dialectology also raise issues of the rapidity with which grammatical structures can shift. It would seem that Romani as spoken in Macedonia is, on the one hand, highly Balkanized and thus raises the issue of macro- versus micro-level characteristics in the definition of the Balkan Sprachbund. At the same time, however, there are basic conservatisms in Balkan Romani that are not found outside the Balkans. While it is not surprising that Balkan dialects of Romani display a higher degree of Balkan features and that dialects whose speakers have been sedentary for a longer period of time should go even further in some respects, as is the case with Arl, especially in Macedonia, it is also noteworthy that the Romani dialects of the Balkans have maintained certain features distinct from the canonical Balkan languages. In this respect, Romani looks more like Greek in relation to the other Balkan languages. The reason for this resemblance, however, is not necessarily the fact that Romani was in significant intimate contact with Greek at an early period. It may have more to do with competition between contact-based influence and social boundary maintenance. We thus have in Balkan Romani a dynamic tension between accommodation and resistance, between influence from multilingualism and codeswitching on the one hand and the maintenance of distinctive norms on the other. The study of Romani in its Balkan context — both diachronic and modern synchronic — thus sheds light on the nature of Balkanisms and differing mechanisms of contact-based interference and resistance.

Notes

1The diaspora Albanian dialects of Thrace (and also Ukraine), however, show the adjective-noun order of their contact languages (see Friedman 1994c).

2Andrej Sobolev of the University of St. Petersburg (XXVII Inter-University Conference on Linguistics and Literature, University of St. Petersburg, Russia, March 1997; PC) has raised the legitimate methodological issue of the Balkan Sprachbund as a consistently definable unit. Despite the original 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 selection of the correspondences used to define the latter have yet to be adequately 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 genetic origin. Contact induced change, by its very nature, involves a complex ecology of choices among competing systems (cf. Muñoz 1996). For the purposes of this paper, it suffices to adduce a number of clearly areal features to support our main point, which is the relevant position of Romani within the Balkan linguistic league.

3While in both genetically and areally related languages shared innovations are crucial in determining the type and degree of relationship, nonetheless, given the fact that contact phenomena involve the selection among competing alternatives, shared archaisms can also be said to have a place in defining the ecology of the linguistic systems. This is especially true when a language in contact is compared with genetically related non-contact languages. Since contact itself is often an impetus for innovation (cf. Trudgill 1992), shared archaisms in a contact situation have a value in areal linguistics that is quite different from their non-diagnostic nature in genetic linguistics.

4Birgit Igle of the University of Sofia (International Conference on Balkan Linguistics: Synchrony and Diachrony, Aristotle University of Thessaloniki, October-November 1997; PC) notes that synthetic cases are still quite strong in Bulgaria. In this, however, as in other matters, the dialects of Macedonia show a greater tendency towards convergence toward a Balkan mean.

5There are also important dialect-specific phenomena, e.g. the -î reported in Sliven Romani (Kostov 1973), which resembles the -îa in many languages of the Arumanian dialect of Beala di supâr [Gorna Belica] (Friedman 1994b). We can also mention here the discourse based type of evidentiality described in Matras (1996). These are, however, beyond the scope of this paper.

References

Demiraj 1994: Demiraj Sh., Gjuhësi Balkanike, Skopje.
Friedman 1989: Friedman V., Toward Defining the Position of Turkisms in Romani, Jezik i kultura Roma, 251-267, Sarajevo.
Friedman 1991: Friedman V., Romani Nominal Inflection: Cases or Postpositions?, Problemy opisu gramatycznego języków słowiańskich, (= Studia gramatyczne 11), 57-64, Warsaw.
Friedman 1994a: Friedman V., Variation and Grammaticalization in the...


Hübchmannová 1996: Hübchmannová M., Derivation of causatives, factitives, syntactic passives and inchoatives in the eastern and western varieties of Slovak Romani and Hungarian Romani, Paper delivered at 3rd International Conference on Romani Linguistics, Charles University, Prague, 5-7 December 1996.


