
100 YEARS OF
Gypsy Studies
❧❧❧❧❧❧

Matt T. Salo, Editor

PUBLICATION NUMBER 5
THE GYPSY LORE SOCIETY

On the Turkish Lexical Component in Romani Dialects and Their
Relationship to Romani Language Planning¹

Victor A. Friedman

University of North Carolina, Chapel Hill

ABSTRACT

Like all languages that have been spoken in the Balkans for over five centuries, Romani has a significant lexical component of Turkish origin. The position of Turkisms in both the Romani dialects and in a Romani literary language is complicated by several factors. Many of these words are ultimately of Persian origin and some may have entered Romani directly from Persian during the earliest period of migration. As such, they are genetically closer to Romani and could be perceived as more desirable than other types of loanwords. Although dialects spoken outside the Balkans have fewer Turkisms, the Balkan dialects are at the forefront of the codification movement. The proportion of Turkisms in Romani is approximately the same as in the colloquial varieties of other Balkan languages. However, while other Balkan literary languages have consciously attempted to eliminate or stylistically lower the position their Turkish lexical elements, Turkish does not occupy the same sociolinguistic and historico-political position vis-à-vis Romani as distinct from other non-Romani languages. All these factors could influence language planning decisions.

The lexical influence of Turkish on Romani reflects the relationship of Turkish to other languages of Europe: In both Romani and other European languages the frequency of Turkish loanwords decreases with an increase in distance from the old boundaries of the Ottoman Empire. Thus, for example, Sampson (1926) does not give a single word of Turkic origin for the dialect of Wales, nor does Bhatia (1963) in his limited vocabulary of a dialect spoken in Philadelphia. Pobožniak (1964) does mention Turkish in his description of Lovari, but only as an intermediary for Armenian and Greek. The presence of a significant lexical component of Turkish origin is a distinctive feature of the so-called Balkan linguistic league, i.e. Greek, Albanian, Macedonian, Bulgarian, Romanian, and Aromanian, and likewise of the so-called Balkan dialects of Romani.² In the formation of the literary languages of the Balkans, however, both historical processes and conscious efforts have resulted in the relegation of many Turkisms to colloquial, expressive, or archaic status. It is here that the question of Turkisms in Romani takes on its most significant aspect: Given that Literary Romani is currently in the process of formation, and that the Balkan dialects are important vehicles of this movement, what is to be the place of Turkisms and how does the current situation compare with that in other Balkan languages? (This relates to the question of the position of foreign loans in general, but Turkisms occupy a special place due to their specifically Balkan nature.)

Studies of Turkish influences in the Balkan languages date from the beginnings of Balkan linguistics (Miklosich 1884, Seliščev 1925), but there has been virtually no significant study of Turkish influence

on Romani. This is a reflection of the unfortunate fact that Romani is generally ignored in studies pertaining to the Balkan linguistic league (but v. Kostov 1973). I have argued elsewhere that Romani, especially Balkan Romani, occupies a significant place in Balkan areal phenomena and requires serious study in this context, as well as in others (Friedman 1986). In the present paper, I shall discuss the relation of Turkisms in other Balkan literary languages to Literary Romani language planning.

Kazazis (1972) gives an excellent survey of the present-day status of Turkisms in the established Balkan literary languages. After briefly discussing the retreat of Turkisms from their strongest position in the nineteenth century, he begins his discussion of their current status with an important section on exceptions to this general retreat. Aside from the fact that many words were not pushed out but merely down (from neutral to colloquial or expressive) or to one side (to archaic or historical), Kazazis points out that as a result of certain historical circumstances some Turkish lexical items actually spread after the fall of the Ottoman Empire. Three cases in point are Serbo-Croatian, Romanian, and Greek. In the case of the first two, the literary prestige of the dialects of regions that were under Ottoman rule for extended periods, Serbia and Muntenia, respectively, led to the spread of some Turkisms into regions that became parts of modern Yugoslavia and Romania where Turkish rule had been short and its influence minimal. In Greek, a new source of Turkisms in the early twentieth century was the exchange of populations between Greece and Turkey, which began in 1923. In Romani, too, it is possible that the role of the Balkan dialects in the progress of the literary language could result in the spread of some lexical items of Turkish origin.

The discussion of Turkish influence on the Balkan languages extends to all manifestations of linguistic structure: phonological, morphological (both inflectional and derivational), syntactic, and semantic. As I have argued elsewhere, in many instances it is impossible to answer definitively the question of whether Turkish served as the source of a given phenomenon, merely influenced the development of a tendency which existed previously, or in fact represents a parallel development without causal connection.

In contradistinction to the situation in phonology, morphology, and syntax, Turkish elements in the lexicons of Balkan languages are readily identifiable as such unless the etymology of a given word is dubious or there has been borrowing back and forth, e.g., if Turkish has borrowed from Greek or Romance and consequently the precise source of the word in other Balkan languages may be uncertain. In the case of Romani this problem of ultimate source and route of entry also contains additional complexities.

In most studies of the lexical influence of Turkish in the Balkans, it is either assumed or specified that no distinction is made between words of Turkic origin and words borrowed by Turkish from Arabic or Persian, since in any case Turkish served as the intermediary by which these words entered the target languages. Romani, however, is in a different position from the rest of the Balkan languages in this respect. Upon leaving India the Roms came into direct contact with Persian and borrowed a number of important words. Some of these Persian words are found throughout the Romani dialects and also in

Turkish, e.g., *baht* 'luck, happiness', others occur in Balkan Romani and Turkish but not in all Romani dialects, e.g. *laf*/*laf* 'word' (English Romani *lav* but elsewhere *vorba*, *svato*, *duma*, *thavali*, etc.) To this can be added the fact that at least some words of Arabic origin could have entered Romani through either a Turkish or Persian intermediary. After eliminating those cases in which clear phonological evidence indicates one or another source, we are left with those words where the precise provenience would be difficult or impossible to establish given the current state of documentation.

The question of the route of Persian borrowings could conceivably influence language planning decisions. In this regard it is instructive to consider the experiences of the other Balkan literary languages in vocabulary building. All of the modern literary languages of the Balkans, except Albanian, had historically and/or modern closely related languages to serve as sources of vocabulary to meet the demands of expanded areas of use. In the formation of each of these languages there were proponents of seeking vocabulary and inspiration from the related languages, the archaizers or purifiers, and there were those who sought to remain close to everyday speech and utilize dialectal potential; they can be called the colloquializers. In the process of vocabulary building, all of the nascent Balkan literary languages, whether archaizing/purifying or colloquializing, made some overt attempts to eliminate vocabulary of Turkish origin, or, in the case of Turkish itself, of Arabo-Persian origin (In many cases the Turkisms of the former are the Arabo-Persianisms of the latter, since Turkish served as the intermediary via which many words of Arabic and Persian origin entered the Balkans. We thus have the interesting situation in which the same term is considered archaic in both Turkey and the Balkans.)

With regard to the elimination of Turkisms in the Balkans, there is a line of reasoning which, although based on nineteenth-century ideas of progress and nationalism, is nonetheless still to be found among those both professionally and nonprofessionally concerned with language. Kazazis (1977) sums up this type of thinking in passage worth quoting. According to the attitude in question:

"... depending on their origin, loanwords differ as to the degree to which they defile a language. Thus, the Romans, the Franks ('(medieval) West Europeans'), the Venetians, all left their linguistic (read: lexical) imprint on Greek. Those were, however, civilized nations, so that their loanwords into Greek are not much of a disgrace and do not wound the "linguistic dignity" of the Greeks as Turkish loanwords do. The latter are a shameful reminder of the centuries-long abject subjugation of the Greek nation to a culturally undistinguished people, the Turks."

In this general context, the position of Romani today can be compared to that of various Balkan languages in the nineteenth century. Until relatively recently it has been used by its speakers almost exclusively in the domains of oral communication. It is divided into many dialects, and while the core vocabulary is more or less uniform (cf. Cortiade 1986), there is a large lexical area beyond this core where vocabulary has yet to be developed or where various dialects have reached individual solutions either by borrowing from different languages or by preserving or creating native forms not

common to the majority of dialects, as in the example with words for 'word' given earlier. One of the tasks facing literary Romani, therefore, is the creation or selection of non-core vocabulary that will be acceptable to and usable by as wide a body of speakers as possible. The situation in Romani with regard to this non-core vocabulary is more complex, perhaps, than it was for the established Balkan literary languages because Romani speakers are spread over a greater geographic area, and vocabulary is correspondingly more divergent.

In the selection and creation of vocabulary, the same basic considerations face Romani as faced the established Balkan literary languages. Romani has a prestigious ancestor -- Sanskrit -- which stands in a direct historical relation to it. Romani likewise has modern related literary languages -- Hindi, Punjabi, Marathi, Gujarati, Kashmiri, etc. -- whose relative degrees of closeness to Romani vary. It has been proposed (Puxon 1978/79) that vocabulary be expanded by means of borrowings from Hindi, and Jusuf attempted to implement this policy in his Romani grammar (Jusuf and Kepeski 1980). The difficulties inherent in such an approach, aside from the obvious dangers of diglossia illustrated by the fate of Greek, can be seen in the need to adapt words to the Romani phonological system and the need to recognize words of non-Indic origin in Hindi, e.g., *zamāni* (Jusuf and Kepeski 1980:211) 'epoch, time', which in fact comes from Arabic *zamān* and probably entered Hindi via Persian. On the other hand, the presence of a loanword in both an Indic language of India and in Romani could be construed as justification for retaining the word in Romani regardless of its ultimate origin or the fact that it entered the languages independently.

This brings us to a consideration peculiar to Romani relative to the other languages being considered here. Indic and Iranian are unique among the groups of Indo-European in being unambiguously descended from a common dialectal ancestor within Indo-European, viz., Indo-Iranian. Unlike the Balto-Slavic and Italo-Celtic dialectal unities, which are still considered moot by some scholars, Indo-Iranian unity is unquestionable. In addition to this fact, Persian enjoys a significant literary prestige and was a language with which Romani came into contact at such an early date that, while it did occur during migration, may have preceded diaspora. Thus while it would make no sense for Bulgarian to have looked to Lithuanian or for Romanian to have looked to Welsh or Irish, the relationship of Romani to Persian is both socioculturally and genetically closer. In terms of emotional attitudes such as those quoted from Kazazis (1977) above, the position of loanwords of Persian origin, regardless of whether or not they entered via a Turkish intermediary, could carry a higher degree of acceptability to some speakers on historical grounds.

By way of a concrete example, let us consider the Turkisms in a specific Romani dialect, that of Agia Varvara in Athens as described by Gordon Messing (1988). According to Messing (1988), the Roms of Agia Varvara came to Greece from Turkey, and some of the older speakers still know Turkish. But the dialect of Agia Varvara has not been subjected to direct Turkish influence since approximately the end of World War I. Given Balkan history, almost the same claim could be made for Albanian, Macedonian, and southern Bulgarian. The dialect as Messing (1988) recorded it thus reflects a situation parallel to that

which we would have seen in other Balkan languages if the literary norms were not yet established.

Messing's glossary contains approximately 1300 main entries. He writes that he restricted Turkisms to what he considered "basic terms" and that he did not include Turkish words that occurred only once in his corpus. While this lexicon cannot, therefore, be taken as complete, (nor was it Messing's intention that it should be) it can be taken as representative. Out of these 1300 main entries, 351 or 27% could be identified as Turkisms or potential Turkisms. (In some cases, the word may actually have entered Romani through some other medium, e.g., directly from Persian as mentioned above or via Slavic as in the case of nouns ending in /-ija/, e.g., *rakiya* 'brandy'. I did not attempt, however, to differentiate these instances due to their mootness or the possibility of secondary influence.) Of this number, approximately half are of native Turkish origin, the other half being mostly Arabo-Persian with about a dozen being ultimately from Greek, Romance, or elsewhere (e.g., *čaj* 'tea' < Chinese). When these Turkisms are compared with those in dictionaries of Turkisms in other Balkan languages, viz, Škaljić (1966) for Serbo-Croatian and Boretzky (1975) for Albanian, it turns out that all but 56 of the Turkisms in the Romani dialect of Agia Varvara are also attested in one or both of these reference works, i.e., almost 85% of the Turkisms in this dialect of Romani are attested in other languages of the Balkans and the figure could be higher if other Balkan languages and dialects were included. A disproportionately small number of these 56 words (12, i.e. a little over 21%) are of Arabo-Persian or other non-Turkic origin. Based on this admittedly limited but nonetheless representative data, it would appear that the Turkish lexical element in Balkan Romani dialects does not differ overwhelmingly from the surface inventory in established Balkan literary languages.

Looking now toward the future of Literary Romani, the following observations can be made with regard to Turkisms. Firstly, while Romani dialects spoken outside the Balkans show far less Turkish influence, the relative progress of Literary Romani in the Balkans may encourage the spread of some Turkish loanwords as in the Serbo-Croatian and Romanian situations mentioned earlier). Secondly, while the conscious attempt to eliminate Turkisms (or Arabo-Persianisms) in the Balkan literary languages has resulted in the stylistic lowering or marginalization of these words, the same motivations need not apply to Literary Romani. Although a tendency towards purism has been manifested, the Roms have fewer nonlinguistic reasons for eliminating Turkisms per se. As has already been mentioned, many of these words are ultimately of Persian origin and therefore more closely related to Romani than to the other Indo-European languages of the Balkans. In fact some were no doubt borrowed directly from Persian during the early history of the Romani people. In addition to this, the elimination of Turkisms from the stylistically unmarked levels of other Balkan languages was in part motivated by political independence from the Ottoman Empire. In the case of Romani, the Turkish language does not occupy a similar political position as distinct from other non-Romani languages. Finally, regardless of the ultimate fate of Turkisms in Literary Romani, their study -- especially in a comparative context -

- will add to our understanding of linguistic history, language contact, and the place of Romani within the Balkan linguistic league.

NOTES

¹An expanded version of this paper with a glossary appeared in French in *Tsiganes: Identité, Evolution* ed. by Patrick Williams (Paris, 1989).

²I am using the term Balkan Romani taxonomically to refer to that group of Romani dialects spoken, for the most part, south of the Jireček line (the boundary between Greek and Roman culture in ancient times, running roughly from the central Albanian coast across northern Macedonia and central Bulgaria to the Black Sea) and generally characterized by an absence or low degree of Romanian lexical influence.

REFERENCES

- Bhatia, R.G. 1963. *A Gypsy Grammar*. Ann Arbor: University Microfilms.
- Boretzky, Norbert. 1975. *Der türkische Einfluss auf das Albanische. Teil 1*. Wiesbaden: Harrassowitz.
- Cortiade, Marcel. 1986. *Romani fonetika thaj lekhipa*. Titograd: Pobijeda.
- Dizikiríkēs G. S. 1975. *Ná Ksetourképsoume té glōssa mas*. Athens: Angyra. (see Kazazis 1977).
- Friedman, Victor. 1986. RomaniTe in a Balkan Context. *Językowe Studia bałkanistyczne*, Vol. 1, 39-48. Wrocław: Polska Akademia Nauk.
- Jusuf, Saip and Krume Kepeski. 1980. *Romani gramatika - Romska gramatika*. Skopje: Naša knjiga.
- Kazazis, Kostas. 1972. "The Status of Turkisms in Present-Day Balkan Languages." *Aspects of the Balkans*, ed. by H. Birnbaus and S. Vryonis, pp. 87-116. The Hague: Mouton.
- _____. 1977. review of Dizikiríkēs 1975. *Folia Slavica* 1,2:301-306.
- Kostov, Kiril. 1973. "Zur Bedeutung des Zigeunerischen für die Erforschung grammatischer Interferenzerscheinungen." *Balkansko ezikoзнание*. 16,2:99-113.
- Messing, Gordon. 1988. *A Greek Romany Glossary: As Spoken in Agia Varvara (Athens)*. Columbus, OH: Slavica
- Miklosich, Franz. 1884. "Die türkischen Elemente in den südost- und osteuropäischen Sprachen." *Denkschriften der Kaiserlichen Akademie der Wissenschaften in Wien, Phil.-Hist. Kl.*, 34, 35, 38. Vienna: Imperial Academy of Sciences.
- Pobožniak, Tadeusz. 1964. *Grammar of the Lovari Dialect*. Kraków: Polish Academy of Sciences.
- Puxon, Gratton. 1978/79. "O Romano Kongreso 2: The Second World Romani Congress." *Roma* 4,2/3:33-104.
- Sampson, John. 1926. *The Dialect of the Gypsies of Wales*. Oxford: Oxford.
- Seliščev, A. 1925. "Des traits linguistiques communs aux langues balkaniques." *Révue des études slaves* 5:38-57.

Škaljić, Abdulah. 1966. *Turcizmi u srpskohrvatskom jeziku*. Sarajevo: Svjetlost.

Victor A. Friedman is Professor of Slavic and Balkan Linguistics at the University of North Carolina, Chapel Hill. He has published extensively in Balkan, Caucasian, and Slavic linguistics, including articles on Romani language planning and Romani verbal categories in the context of the Balkan linguistic league. He served as president of the Gypsy Lore Society, North American Chapter 1984-86.