

Dialect variation in Romani and its implications for Romani historical linguistics

Vijay John

University of Texas at Austin

1. Introduction

Romani is a language of Indo-Iranian origin, traditionally classified as Central Indo-Aryan. It is spoken by the Romani people (a.k.a. “Gypsies”), who are of South Asian descent but live primarily in Europe and the Americas. The Romani people are dispersed throughout these continents and have lived there for hundreds of years, so there are many

dialects of Romani, especially in Europe. Previous work on Romani historical linguistics, e.g. Turner (1926), has paid little attention to the variation in Romani dialects. However, we certainly cannot reconstruct Romani's South Asian ancestor without knowing how these dialects split up, so we need to gradually reconstruct back from as many varieties as possible. Furthermore, dialect variation can account for what may initially appear to be regular sound changes shared with the Indo-Iranian languages of South Asia. At the same time, there are also situations where dialect variation complicates what we know about Romani.

2. Dialect variation in Romani

Romani can be broadly classified into three dialect groups (Matras, 2000), which appear to have been the result of three separate migrations into Europe from present-day Turkey. The three dialect groups are Northern (a.k.a. Stratum I), Balkan (Stratum II), and Vlax (Stratum III). Northern Romani includes most varieties of Romani; one example of a Northern Romani variety is Welsh Romani, spoken in Wales. This stratum may represent the earliest migration out of Turkey. A second migration is represented by the Balkan varieties, spoken throughout the Balkans and including Gurbeti, which is spoken from Bosnia and Herzegovina to Bulgaria. Vlax Romani or Stratum III is the result of a migration at approximately the time of the Ottoman invasion of Europe; this stratum apparently broke up only after a period of common development in Romania. Examples include the Kalderash variety of Romania, which is now spoken elsewhere in Europe and the Americas as well, and the Mačvànò variety, originally spoken in Serbia but now almost exclusively spoken in the Americas. In addition, a European language may be combined with a Romani lexicon. For example, Fennoromani is essentially Finnish with a large, though limited, Romani vocabulary.

The variation between these dialects is very often due to substantial contact with local varieties of European languages. Some of the variation is fairly straightforward and can arise as a result of either independent development or language contact. For example, palatalization before a high front vowel is very common in Romani (Matras, 2000). While the second person masculine singular possessive pronoun (i.e. 'your (masc. sg.)') in most varieties is *tiro*, it is *čiro* in some varieties and *kiro* in others. This seems to be the result of an independent development, but palatalization can also occur as a result of language contact. Fennoromani shows palatalization of velars, e.g. *čher* 'house' < *kher* (cf. *kher* in other varieties of Romani such as Kalderash) due to influence from Swedish, which has the same process. One other example of straightforward variation due to language contact is the adoption of the voiceless lateral fricative and devoiced alveolar trill from Welsh into Welsh Romani.

There are other cases where dialect variation complicates the historical picture. In some of these cases, language contact and dialect variation can cause changes to appear more regular than they really are. In others, the source of the dialect variation may be difficult to find or explain, although it is likely that language contact played an important role.

3. How dialect variation explains some apparent sound changes

There are some sound changes that may at first glance seem to occur regularly in Romani. It is claimed that some of these are even shared with Dardic languages. In fact, it is more likely in these cases that this is the result of dialect variation rather than regular phonological change. Four examples are discussed below.

One sound change that Turner (1926) proposed that Romani shared with three Indo-Iranian languages of South Asia (Kashmiri, Sindhi, and Lahnda) was as follows: Word-medial *r* preceded or followed by a consonant was metathesized to the position after the first consonant of the word, provided that this first consonant was dental (or perhaps labial). For example, Sanskrit **tarṣáyati* ‘he/she/it fears’ > Romani *trašel*. However, metathesis is very common in the languages of the world, and usually it occurs only sporadically, not as the result of a regular sound change. The same seems to be true in Romani; there are other words where metathesis occurs sporadically, and there seem to be exceptions to this rule, e.g. Sanskrit *tárdati* ‘splits open, sets free, hurts, kills’ > Welsh Romani *tardel* ‘drags’ without the expected metathesis (Turner, 2006). Thus, metathesis is most likely subject to dialect variation in Romani.

Another sound change that Zoller (2001) suggests Romani shares with Dardic languages is what he calls “metathesis of aspiration.” He says that in these languages, when an aspirated stop occurs word-medially, the aspiration moves to the first consonant. For example, the Sanskrit word for ‘armpit’ is *kákṣa*, and the equivalent in Romani that he lists *khak*. Again, though, metathesis is often a sporadic change, and there are varieties of Romani that have *kakh* for ‘armpit’ as well.

One other apparent sound change that e.g. Matras (2000) has pointed out is that *kh* often changes to *x* word-initially in Romani, e.g. Sanskrit *khā’dati* ‘he/she/it chews, bites, eats’ > Romani *xal* ‘he/she/it eats’. No one has explained why this change may have taken

place, although Romani could have adopted this sound from Iranian languages. This change, too, happens only sporadically in some varieties of Romani. Bosnian Gurbeti has *khuš-* and *xuš-*, which both mean ‘to whistle’ (Boretzky and Igla, 1994), so *kh* > *x* / #_ is certainly not a regular change that occurs in all dialects of Romani.

Finally, in the Romani words *jekh* ‘one’ and *nakh* ‘nose’, the final stop is aspirated even though it was unaspirated in Sanskrit, as it still is in most Indo-Aryan languages (cf. Hindi/Urdu *ek* ‘one’ and *nāk* ‘nose’). In Kashmiri, though, voiceless stops are frequently aspirated word-finally, so the Kashmiri words [ak^h] ‘one’ and [na:k^h] ‘nose’ both have aspiration just like their Romani equivalents. At first glance, this may suggest that Romani and Kashmiri both share word-final aspiration, but in fact, aspirated stops also occur sporadically in Romani. The word for ‘nobody’ in most varieties of Romani is *kònik*, but Kalderash and Gurbeti have both *kònik* and *khònik* (Boretzky and Igla, 1994). This is probably due to contact with European languages, which mostly do not contrast aspirated and unaspirated stops. The sporadic occurrence of aspirated stops may also explain why metathesis of aspiration seems to occur in Romani.

Thus, metathesis of *r* and of aspiration, word-initial *kh* > *x*, and word-final aspiration are all more easily explained as the result of dialect variation and sporadic sound changes, even though they may appear to be regular sound changes. In fact, some may appear to suggest a connection to specific Indo-Iranian languages of South Asia, but they may be more easily explained as the result of independent innovation and contact with European languages. Unfortunately, in other situations, dialect variation is more difficult to explain.

4. Problematic dialect variation

So far, we have seen that dialect variation can be unproblematic for historical linguistics and that it can offer a different analysis of some apparent sound changes in Romani. There are also cases where dialect variation is problematic for historical linguistics and cannot be explained easily. One problematic case is that different varieties of Romani have different sets of rhotics, and explaining the origin of these rhotics is difficult. Another case, to be discussed later, is that Mačvànò Vlax has two very distinctive affricates that require a historical explanation.

Some varieties of Romani have two rhotics, and others have only one. Those varieties that have two rhotics do not all have the same sounds (Matras, 2000). In some varieties,

one rhotic is an alveolar tap (sometimes phonetically realized as a trill by some speakers), and the other is a voiced uvular fricative or trill. We may represent the alveolar tap graphemically as *r* and the uvular rhotic as *rr*, because in other varieties, these rhotics (especially *rr*) may correspond to different sounds, and this allows us to conveniently indicate which sounds in one dialect correspond to which in another. In some other varieties, *r* is an alveolar tap, while *rr* is an alveolar trill. In Bulgarian Gurbeti, *r* is again an alveolar tap, but surprisingly enough, *rr* is a retroflex lateral approximant. The origin of these rhotics is unclear and subject to debate. Many believe that *rr* comes from retroflex stops in Indo-Iranian words, but Hancock (2006) points out that this does not explain why *rr* sometimes occurs in loanwords. In addition, there are words that have *r* corresponding to retroflex stops in their Indo-Iranian cognates. Surrounding varieties of European languages seem to play a role here; in particular, nonstandard varieties of Romanian help to explain why varieties in Romania have a uvular rhotic, and some varieties of other European languages may also help to explain the presence of a retroflex lateral. Standard Romanian supposedly has only one rhotic, which is an alveolar trill. However, nonstandard varieties of Romanian have a uvular rhotic contrasting with an alveolar one, like some varieties of Romani (Bardu, 2007). Even some speakers of Standard Romanian use a uvular rhotic instead of an alveolar trill. Similarly, descriptions of phonetic inventories in European languages do not usually include a retroflex lateral. Yet Moosmüller (2010), for instance, reports that Viennese German does have a retroflex lateral as an allophone of /l/. Thus, perhaps the retroflex lateral is not extremely rare in Europe and may have been acquired due to contact with European languages, rather than preserved from India (where many languages have a retroflex lateral approximant).

Rhotics are a well-known problem in the historical linguistics of Romani. Another problem that has received less attention is specific to one dialect; this problem is affricates in some varieties of Romani, especially Mačvào Vlax. Some varieties of Romani have three affricates: *č* [tʃ], *čh* [tʃ^h], and *dž* [dʒ]; these are probably the original pronunciation preserved from India, since most Indo-Iranian languages have these same three affricates. However, in some other varieties, such as Kalderash, *čh* is pronounced [ʃ], and *dž* is [z]. In Mačvào Vlax, these are sequences of a retroflex stop followed by [r], i.e. *čh* [tʃr] and *dž* [dʒr] (Hancock, 1995). Perhaps the following chain of sound changes took place in all of these varieties of Romani: [tʃ^h], [dʒ] > [tʃ], [dʒ] > [tʃ], [dʒ]. That is to say, the aspiration from [tʃ^h] was lost and the fricative portion made retroflex in both affricates, and then the stop portion of each affricate assimilated to the place of articulation of the fricative portion. After this, [tʃ], [dʒ] > [tʃ], [dʒ] in Mačvào Vlax and to [ʃ], [z] in Kalderash and some other varieties of Romani. In this case, too, the retroflex sounds may have come from contact with European languages rather than from Indian languages. Retroflex fricatives occur in Polish, for example, and retroflex stops certainly occur in Swedish, so it is possible that nonstandard varieties of other European languages also have these retroflex sounds.

In both of these cases, dialect variation poses a problem for historical linguistics. If different varieties of Romani have different rhotics, then there is a problem of figuring out whether there was originally one rhotic that split in some varieties (and evolved differently in different regions), or whether there were originally two that merged in the other varieties. This problem remains unsolved. Similarly, there is a problem of figuring out how affricates evolved in some varieties of Romani that do not have what are believed to have been the original affricates. In the case of these affricates, we can at least propose a set of sound changes that may have led to the variation that we see today. Substantial contact with European languages and nonstandard varieties of those languages seems to have played a role in the development of the sounds we see in modern varieties of Romani, and the development of uvular rhotics and retroflex sounds may well be a result of this heavy contact.

There is one more problem for Romani historical linguistics, having to do with the classification of Romani into three strata. The idea that the Romani people entered Europe in at least three separate migrations from Turkey suggests that each stratum represents a different stage of one variety spoken in Turkey. It is possible that the historical picture is even more complicated than this (perhaps there were already several varieties were spoken in Turkey). Nevertheless, this implies that a full understanding of Romani dialect variation (from a diachronic perspective) requires a reconstruction of at least one variety of Romani spoken in Turkey. If all three strata are descended from one variety spoken in Turkey, then each stratum represents a different stage in the evolution of that variety, so it is necessary to understand how that variety evolved over time.

5. Conclusion

It is impossible to reconstruct the ancestor(s) of Romani without taking dialect variation into account, despite the lack of attention paid to this dialect variation in the context of Romani historical linguistics. Sometimes this variation is unproblematic, but at other times, it is difficult to account for and needs further research in order to be explained. What appears at first to be a regular sound change may turn out to be a change that happens idiosyncratically only in some varieties of Romani. Very often, this variation is due to long periods of contact with varieties (not necessarily standard varieties) of European languages.

References

- Bardu, N. (2007). "Among the Aromanians in Grabova (Greava), Albania. Sociolinguistic Observations." *Annals of Ovidius University Constanta – Philology* 18, 17-28.
Constanța, Romania: Ovidius University Press.
- Boretzky, N. and Igla, B. (1994). *Wörterbuch Romani-Deutsch-Englisch für den südosteuropäischen Raum: Mit eine Grammatik der Dialektvarianten*. Wiesbaden, Germany: Harrassowitz Verlag.
- Hancock, I. (1995). *A Handbook of Vlax Romani*. Columbus, OH: Slavica Publishers, Inc.
- Hancock, I. (2006). "On Romani Origins and Identity." Austin, TX: Romani Archives and Documentation Center (RADOC).
- Matras, Y. (2000). *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Moosmüller, S. (2010). "The roles of stereotypes, phonetic knowledge, and phonological knowledge in the evaluation of dialect authenticity." *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication."*
- Turner, R. L. (1926). "The Position of Romani in Indo-Aryan." *Journal of the Gypsy Lore Society*, 3rd series, 4(4). Liverpool: Gypsy Lore Society.
- Turner, R. L. (2006). *A Comparative Dictionary of the Indo-Aryan Languages* (3rd ed.). London: Oxford University Press. Retrieved from <http://dsal.uchicago.edu/dictionaries/soas/>.
- Zoller, C. P. (2010). "Aspects of the Early History of Romani." *Acta Orientalia* 71, 243-312. Oslo: Hermes.