On the Koïné Origins of Romani: A Study of Romani Words and Numbers of Indic Origin

Vijay John

November 2008

Abstract

Modern research in or related to Romani Studies generally supports the idea that the Dom (a non-Romani Indian diasporic people in the Middle East) left India long before the Romani people and possibly before the Lom (another Indian diasporic people in Armenia) as well. Much evidence supports the modern idea that Romani originated as a military koïné. The cognates of many Romani numbers and synonyms that can be found in a variety of Indian languages form a significant part of this evidence.

1. Introduction

Scholars have been debating the origins of the Romani people and language for centuries. With the revelation in the 18th century that the Romanies are of Indian origin, several views arose as to where exactly they came from and how they arrived in Europe. The earliest view, originally proposed by Heinrich Grellman and others in the 18th century, was that the Romanies' ancestors in India were originally *Śudras*, low-caste people who performed menial jobs; this was also the proposed explanation as to why Romanies

1

undertook similar professions in Europe. European scholars in Romani Studies later learned that another diasporic Indian people calling themselves *Dom* lived in the Middle East, that yet another diasporic people called *Lom* lived in Armenia, and that both seemed to have similar occupations to those of the Romani people.[1] It was also found that Ferdowsi's *Shah-Nameh* includes a story in which approximately 10,000 *Luri* musicians from India were sent west to present-day Iran; James Harriott, a British army officer in India, suggested that they later moved further west and that these musicians were thus the ancestors of the Romani people. The scholars came to the conclusion that all three peoples were part of an original migration out of India, but the *Dom-Lom-R(r)om* split, as it was called, occurred only after the migrating Indian people reached Iran.

John Sampson proposed a detailed argument for this split, which is still widely supported today. In fact, many researchers in Romani Studies still accept the view that the *Dom*, *Lom*, and *R(r)om* are all descendants of *Śudras* who migrated out of India, then separated into three waves after arriving in Iran. More recently, however, various researchers have suggested an alternative origin. Ralph Turner, for example, has proposed that the Romanies' ancestors did not split, as originally believed, into three groups outside of India. Rather, the *Dom*'s ancestors left India in a migration separate from that of the ancestors of the *Lom* and *R(r)om*. The latter group may indeed have split into two groups, eventually resulting in the formation of Lomavren, i.e. a lexicon of Indic words retained by the Lom, and the Romani language. Nevertheless, only in Armenia can we say for certain that a split occurred (assuming that they were together to begin with). In a much

more recent work, Adrian Marsh (2008) provides a variety of historically based arguments to support Turner's position.

In addition, current research points to a departure from India no earlier than the 11th century, and there is a very strong likelihood that Romani, like modern Hindustani (see appendix), arose as a military koïné. This can be demonstrated through the linguistic nature of Romani, as exemplified below. This paper's main task, however, is to analyze words and numbers in Romani that come from various Indic sources and present them as evidence that Romani was originally a koïné. It must be noted that the analysis below is based on the regions in which the Indian languages in question are currently spoken and not on where these languages were spoken at the time of the Romani exodus.

2. Survey

The Romani words and numbers are primarily influenced by three subgroups of the Indo-Iranian languages:

1. Central Indian (a.k.a. *Śaurseni*) languages, nowadays mainly spoken between Gujarat and Rajasthan (northwestern India) and Uttar Pradesh (between Delhi and Nepal);

2. Dardic and Nuristani languages, currently spoken in Kashmir and Nuristan; and

3. Iranian languages.

Some similarities with the mediate (*Ardhamāgādhi*) and Eastern Indian (*Māgādhi*) languages also exist in Romani. The former is spoken in Bihar (southeast of Uttar

Pradesh), as well as in some nearby parts of other Indian states (e.g. in southeastern Uttar Pradesh). The latter is spoken in northeastern India, Bihar, and along the eastern coastline, as far south as the area directly north of Hyderabad. However, there are no Romani numbers that are similar only to these two groups; all numbers of Indic origin can be traced to the three main groups mentioned above (Central Indian, Dardic, and Iranian).

In some cases, the Pahari languages, now spoken in the Himalayas northwest of (and within) Nepal, also seem to have influenced Romani numbers. This is not unlikely; according to Indian geneticist Vijender Bhalla, Romanies are related to "Jat Sikhs [and] P[u]njabi Hindus" as well as Rajputs, and the area where the Pahari languages are spoken is very close to Delhi and Punjab (Hancock 2003:9).

The above description locates the regions where these languages are primarily spoken in the present day. However, it is known that the Dardic languages, though now spoken in the "Northwest" (i.e. the regions of Kashmir and Nuristan in the north of India, Pakistan, and Afghanistan), were originally spoken further south. Thus, some of the Indian ancestors of the Romani people, we may assume, spoke at least one Dardic language, even if they were not originally from the Northwest or were not native-speakers.

Not all the words discussed here originate in any single language belonging to any one of the three groups listed above. Most of the words mentioned in this paper are synonyms; in Romani, there are many synonyms of Indic origin that can be traced back to different

Sanskrit (and sometimes different Middle Indo-Aryan, or MIA) roots. These synonyms seem to have relatively unclear origins, since they tend to have cognates in a wide variety of New Indo-Aryan (NIA) languages as well. By contrast, the numbers often have rather distinct origins; though they are related to their equivalents in many NIA languages, evidence is available as to which languages have the closest cognates to them.

3. Evidence refuting older theories

Turner proposed that the Domari people did not have the same linguistic origin as the Lom and Romani peoples. It was previously argued that the Domari people split off from the original migration after passing through Iran. One piece of evidence that supports Turner's position is the proportion of words of Persian origin shared with Romani: if the Domari and Romani peoples were together until they had passed through Iran, their languages should share Persian loanwords. But while 50% of Romani words of Persian origin share the same origin as their equivalents in Hindustani, only 16% of Romani words of Persian origin share origins with their equivalents in Domari (Hancock 2006:9).

Many arguments have also provided evidence that Romani originated in a military koïné. It has already been argued that Romani and Hindustani both classify nouns that were originally neuter with the same gender (e.g. [pã:ni:] "water" is masculine in both languages); more than half of Romani's loanwords from Persian are present in Hindustani as well. Both of these facts are evidence that the Romanies' ancestors left India after the Ghaznavid invasion. In addition, Romani has many military words of Indic origin, e.g. *bust* "spear" and the verb stems *lod-* "to set up camp" and *rad-* "to strike camp."

The mixed nature of the Romani lexicon is illustrated, for example, by numbers of Indic origin, which represent various Indo-Aryan groups. Thus, the numbers for "two" are of Indo-Aryan origin, whereas the number for "four" is of Dardic origin and "five" is of mixed origin. This can be extended to other words as well.

4. Romani numbers of Indic origin

Those numbers in Romani that are of Indic origin provide evidence for the formation of a koïné. The other numbers in Romani are analytical numbers, some of which have no Greek origin as they are combinations of the Indic numbers, i.e. one through six and ten, twenty, or a hundred. The numbers 11-16 combine the Indic numbers between one and ten using the infix *-u-* (adopted from Persian/Farsi *-o-* meaning "and"). Numbers above 20 (i.e. 21-26, 31-36, etc.) instead use another infix *-ta-*, derived from Old Indo-Aryan *athha*. For example, in Vlax Romani (see appendix), "eleven" is *deš-u-jekh* and includes the Persian-derived infix *-u-* while "twenty-five" is *biš-ta-pandž* and includes the Indic infix *-ta-*.

Some speakers of Vlax Romani use numbers of purely Greek origin for 30, 40, and 50 (*trànda*, *sarànda*, and *pìnda*, respectively). However, the numbers 30, 40, 50, and 60 can also be expressed using Indic words for e.g. "three times ten"—*trivàrdeš*[2]. In addition, tens beginning with an even number may be expressed in terms of twenty, e.g. "two times twenty" for forty. Thus, 80 may be either *oxtovàrdeš* ("eight-times-ten") or *štarvàrbiš* ("four-times-twenty").

The remaining numbers are either of Greek origin or of mixed Indic and Greek origin. The numbers seven through nine, for example, are adopted from Greek: the Romani numbers *efta*, *oxto*, and *inja* are pronounced (almost) exactly like their equivalents in Modern Greek. The number nineteen in Vlax is *deš-(u)-inja*, a combination of Indic *deš* and Greek *inja*; the infix *-u-* may be dropped when *deš* is combined with the three Greek numbers in Romani (Hancock 1995:164-66).

Some of the Romani numbers also reflect phonological rules of Indo-Iranian languages, including Sanskrit and Pahari languages. In fact, the first number *jekh* (which is interchangeable with *ekh*) reflects a rule that occurs not only in some Indian languages (including at least one dialect of Hindustani), but also in various languages around the world, e.g. Slavic languages (especially Russian) and even (though to a lesser extent) Korean and Japanese. This and several other phonological rules will be discussed in subsequent sections in greater detail.

A wide variety of Indo-Iranian languages spoken south of Kashmir appear to have influenced the Romani numbers. The names of the Indic numbers are especially similar to their equivalents in the Western Pahari languages and in Hindustani. Gujarati and Punjabi do not appear to have influenced these words quite as much.

Sanskrit itself seems to have contributed somewhat to the formation of the Romani numbers. It seems as if many of the sound changes from Sanskrit to Romani are also

found in Dardic languages. To a lesser extent, they may also be found in the Central Indian languages or Hindustani or both. It is also possible that the influence of Sanskrit was at least partly due to Rajputs who could speak Sanskrit; even in today's India, many upper-caste Hindus speak Sanskrit (some even claim it as their first language!), and it was not uncommon for Kshatriyas (members of the warrior or ruler caste with which Rajputs are usually associated) to marry Brahmins (from the caste of scholars and priests)[3] who knew Sanskrit.

The Romani language has a considerable number of loanwords from the Dardic languages, which also have influenced and contributed numbers that exist in modern Romani. The numbers in Dardic languages are almost never identical to their equivalents in Romani, with one possible exception: The number "twenty"—*biš* in Vlax Romani— might be pronounced just like its equivalent in the Kohistani languages, *bīš*. Only a very few Western Pahari languages have words for twenty that sound as similar to *biš* as the Kohistani languages (Grierson 1968a:23).

5. Numbers of Central Indian origin

The numbers *jek* (one, in some dialects of Romani), *duj*, *do* (two), and *trin* (three) are most closely related to Indian (non-Dardic) Indo-Iranian languages. Nevertheless, each one has considerably different origins; each of these numbers is most closely related to equivalents in the Indo-Iranian languages of its own region of the Indian Subcontinent.

Jek is comparable to its equivalent in Hindustani and many other Indian languages (including some Dardic languages), *ek*. The addition of the initial [j] may be due to influence from the Dardic languages, in which a similar sound change has occurred. It is consistent with the addition of initial [j] in such words as *jakh* "eye" (compare Hindustani [ã:kʰ]) and *jag* "fire" (Hi. [a:g]). Like Romani, Dardic languages differ with respect to the addition of [j] in the word for "one." For example, the word for "one" in Poguli is *yakh*, but in Kashmiri it is *akh* (or *okʹʹ*). Note that Grierson (1968a:3) lists *yēk* as the equivalent for "one" in the Banaaphari dialect of Hindustani and the Jaipuri dialect of Rajasthani, as well as *yek* (among other words) in Khaskuraa or Eastern Pahari.

For "two," *duj* or some very similar form (e.g. *dūī*) is used mainly in eastern India: as shown in figure 1 below, it is used in Bengali, Oriya, Assamese, and some Bihari languages. It is also, however, found in some languages considered dialects of Hindustani (e.g. Chhattisgarhi and Kanauji), as well as Nimadi (a Rajasthani language spoken primarily in the modern-day Indian states of Madhya Pradesh, Maharashtra, and Uttar Pradesh). *Duj* is not limited to these areas of Northern India; many Pahari languages, and even the Doda Siraji dialect of Kashmiri (sometimes known as Siraji of Doda), use *duj* (Grierson 1968a:4).
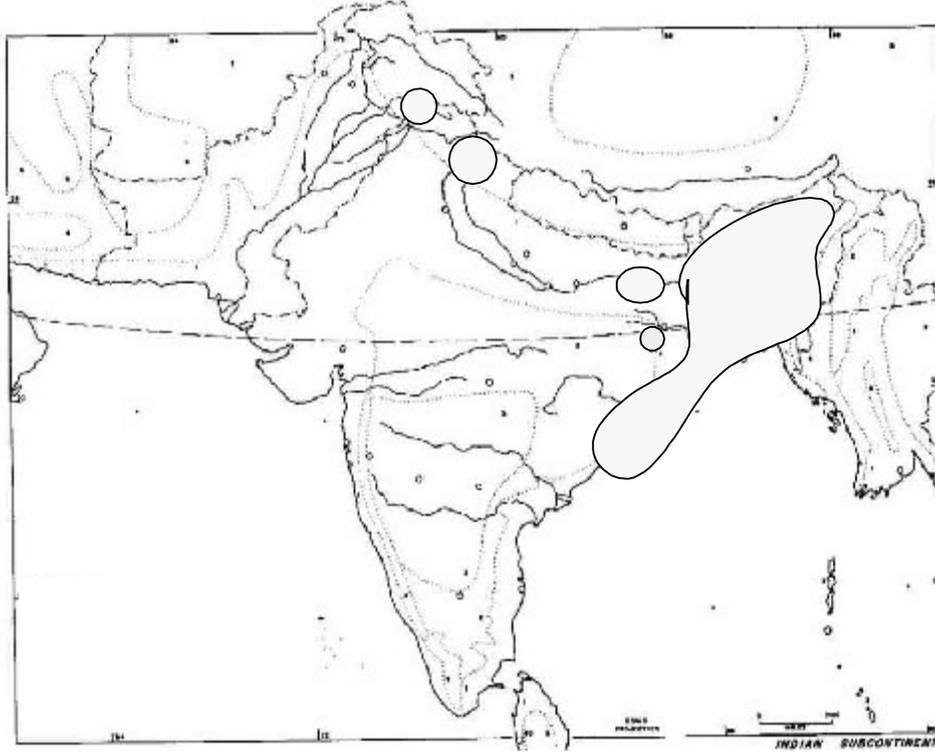
Figure 1. Indian regions with "two" as *duj* or *dūī*

In Romani, *duj* is the nominative form of the word "two"; *do*, however, is used in the oblique case. *Do* is also the Hindustani word for "two" and is used in many Indo-Iranian languages, especially those spoken in the region where Hindustani is the *lingua franca*. In Hindustani, unlike Romani, *do* is the only word for "two"; there are no special words for numbers in the oblique case. In some Dardic languages, though, both nominative and oblique forms exist for the numbers.

Both Shina and Kashmiri include oblique as well as nominative forms for the numbers. In Kashmiri, the nominative form of the number "two" is *zi(h)*, but there are two oblique

cases; in Oblique I, "two" is *dvan*, while in Oblique II (the Agentive case) it is *dvaw* (Edelman 1983:309).

Grierson (1968a:4) lists two Indian languages with both *duj* and *do*. One is Chhattisgarhi, now the official language of the state of Chhattisgarh, formerly a part of eastern Madhya Pradesh. The other is Pangwali, a Pahari language spoken in present-day Himachal Pradesh as well as Kashmir. It is possible that Pangwali actually uses both *duj* and *do*, but Chhattisgarhi probably does not have *do* and certainly does not use it as it is used in Romani.

Grierson (1968b:185-236) provides no evidence that *do* is ever used in Chhattisgarhi. In fact, it may have been a mere misprint; in Volume VI, in which Chhattisgarhi is described in more detail, only *duj* is listed as a translation of the number "two." Furthermore, there are many Hindi-language grammars of Chhattisgarhi, and none of them seem to list *do* as a number in the language (though it is indeed a number in Hindi). Jain (1984:121), for example, lists [ɖu:], [ɖuwi:], [ɖu:j], [ɖu:wo], and [ɖu:go], but not [ɖo]. In any case, Romani reserves *do* for the oblique case—a case that Chhattisgarhi lacks. It is therefore probable that Chhattisgarhi does not even include both *duj* and *do*.

More written evidence is available for Chhattisgarhi than for Pangwali. The former is spoken by a larger population; Pangwali, on the other hand, is mainly concentrated in remote areas of the Himalayas. In Vol. IX Part IV, Grierson (1968c:859) does indeed present *do* in context in Pangwali (in a translation of a Biblical parable) without

providing a text including *duj*. He also include a description of a Kashmiri-based sound change that may even explain the presence of both numbers in Pangwali:

> In Kāshmīrī, epenthesis is a common feature of the language. Thus the base *kōr-*, a girl, becomes *kūrⁱ*, when the nominative termination is added. The same occurs in regard to the same word in Pa[n]gwā[l]ī. Here we have the word *kōā* (i.e. *kō[R]ā*), a boy. But when the word is made feminine by changing the final *ā* to *ī*, we get *kūī* or *kū[R]ī*, a girl [Grierson 1968c:848].

Perhaps a similar vowel change occurs so that *do* is masculine and *dūī* is feminine. If so, Proto-Romani may have had both [ḍo] and [ḍuj], contrasted by gender, before the ancestors of the Romani people made the exodus out of India. If not, Proto-Romani may have used both numbers based on languages (especially Dardic languages) that use two similar words for "two" (e.g. *do* and *du* in Ashkun, a Nuristani language), and its use of separate forms for nominative and oblique cases is probably adopted from Dardic languages.

As far as Modern Indian languages are concerned, *trin* is most similar to the Standard Gujarati word for "three," pronounced [ʈrʌɳ] (Dave 1995:30). However, the Sanskrit word is [ʈrĩːɳiː][4], and as Turner (1924:40) notes, this is one of the many words in Romani in which [r] after a dental has been preserved. Thus, the Proto-Romani form was probably identical to its current form. The absence of the final [iː] is consistent with several Indian languages, e.g. Hindustani ([ʈĩːn]), in which the final [iː] was also dropped. It is most likely that the Romani word for three comes from Sanskrit. Still, considering that

(a) there seem to be no other NIA languages that include both the [r] and some nasal consonant, and that

(b) the Lomavren word for "three" is *tran*,

either the Romani word may have been influenced by Gujarati as well, or Romani,

Lomavren, and Gujarati may have maintained both the [r] (immediately following the

voiceless dental stop) and [n]. In particular, the Lomavren word may have been
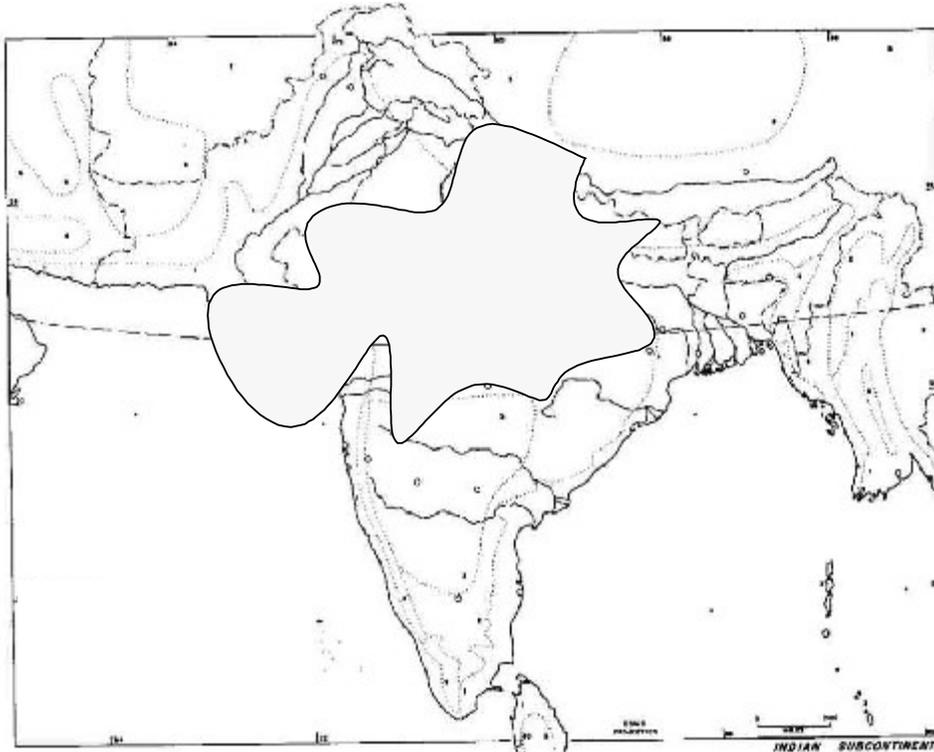
influenced by Gujarati.

Figure 2. Regions that may have influenced the Romani word *trin* "three"

Central Indian (and even Mediate) languages besides Gujarati may have also contributed

to the formation of the word *trin*. For this reason, figure 2 includes not only Gujarat but

also the Hindustani-speaking regions of India (and Pakistan) as areas in which languages

with similar words for "three" are spoken.

All of these numbers seem to have very different origins, but they are related to their equivalents in Central Indian languages. Not all of the other Romani numbers, however, can be correctly identified as strictly Central Indian in origin.

6. Two Dardic numbers

The Romani numbers *štar* "four" and *šov* "six" are both of purely Dardic origin. These are the only one-digit numbers in Romani to begin with the sibilant *š*. Similarly, in Hindustani, these are the only one-digit numbers that begin with the affricates [tʃ] or [tʃʰ]. Whether this means a cause-and-effect relationship (i.e. Indian languages beginning only two numbers with sibilants) is not known at this point. However, this presents one reason to believe that the current numbers differ little from their equivalents in Proto-Romani.

Štar is most closely related to *štā*, its equivalent in Wai-ala, a Dardic language spoken in the southeastern part of the Afghan region Nuristan. The only other Indo-Iranian language listed in Grierson (1968a:9) whose word for "four" begins with the compound *št-* is Bashgali or Kati, another Nuristani Dardic language, which uses a slightly different word *što*. Since the Sanskrit word for "four" is *catvārah* (masculine), and most Indo-Iranian languages include an *r* in their equivalents, probably *što* and *štā* are derived from earlier forms including an *r*, and Romani *štar* may be derived from one of these earlier forms. However, because Romani is a contact language, it could also be argued that the formation of *štar* may have been a result of "reinforcement by convergence." In other words, the word-final *-r* was kept in Romani at least partly because Indo-Aryan languages retain it, but the consonant cluster *št-* was derived from a Dardic or Nuristani

language. The two languages that have word-initial *št-* have no *r* in their words for "four," whereas most other languages include *r*, so the Romani form could partly be a result of blending.
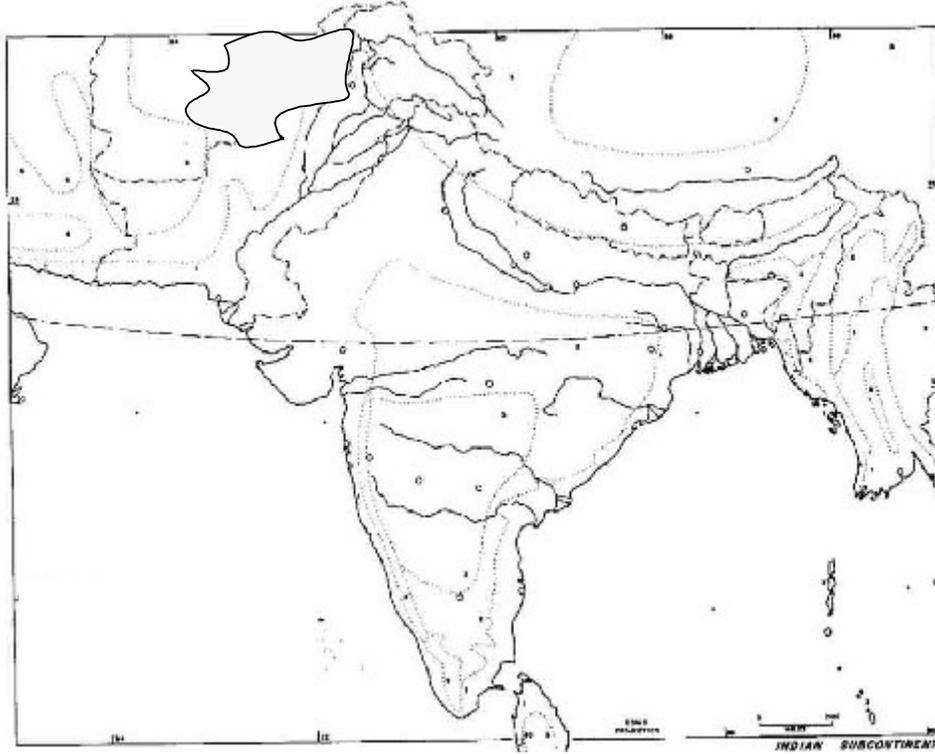

Figure 3: Region with "four" as *što* or *šta*

The region represented in Figure 3 with vertical lines is Nuristan (i.e. northeastern Afghanistan), in which both Bashgali and Wai-ala are spoken.

In *The Piśāca Languages of North-Western India*, Grierson (1969:15-18) explains how the Dardic cognates of *štar* are ultimately derived from its Sanskrit equivalent *catvārah*. He accounts first for the disappearance of the first *a*:

> Aphæresis and syncope both occur. The latter is most common in B[ašgalī], where it gives rise to some curious combinations of consonants at the beginning of a word [Grierson 1969:15].

Soon afterwards, he also explains the simplification of the word-initial affricate to a fricative:

> The letter *c* is sometimes changed to the corresponding affricata *č*, and in Kāfir is liable to be further weakened to *š* and *ž* [Grierson 1969:18].

The "Kāfir, or Western, Group" to which Grierson refers includes both Bashgali and Wai-ala. Romani itself may have adopted these sound changes, so that the current word for "four" is nearly identical to the number used when leaving Kashmir.

If only a few Dardic languages have cognates of *štar*, then many more use words for "six" similar to the Romani equivalent: *šov* (see figure 4 below). The Romani number for "six" closely resembles the words for six in Bashgali (*šo*), Kalasha (*šōh*), Gawar-bati (*šoh*), and the Kohistani languages (*šō* in Garwi and Torwali, *šoh* in Maiyã a.k.a. Indus Kohistani) (Grierson 1968a:13). It is possible that the *v* was simply added to the original Dardic-based word for "six" as a phonetic change from final Dardic [o] to final Romani [oʊ]. In his *Comparative Dictionary of the Indo-Aryan Languages*, Turner (1969:743) gives "six" in the Kohistani dialect as ṣvā and in the Kolaa dialect as ṣuā, so this is another possible source of Romani [oʊ]. Another possibility is that the word-final [ʊ] was simply a consequence of the tendency in Romani to sometimes add [ʊ] to words ending in [o]; this also occurs in the word *vov* "he" (compare Hindustani *woh* [ʊo] "genderless 3rd person singular pronoun").
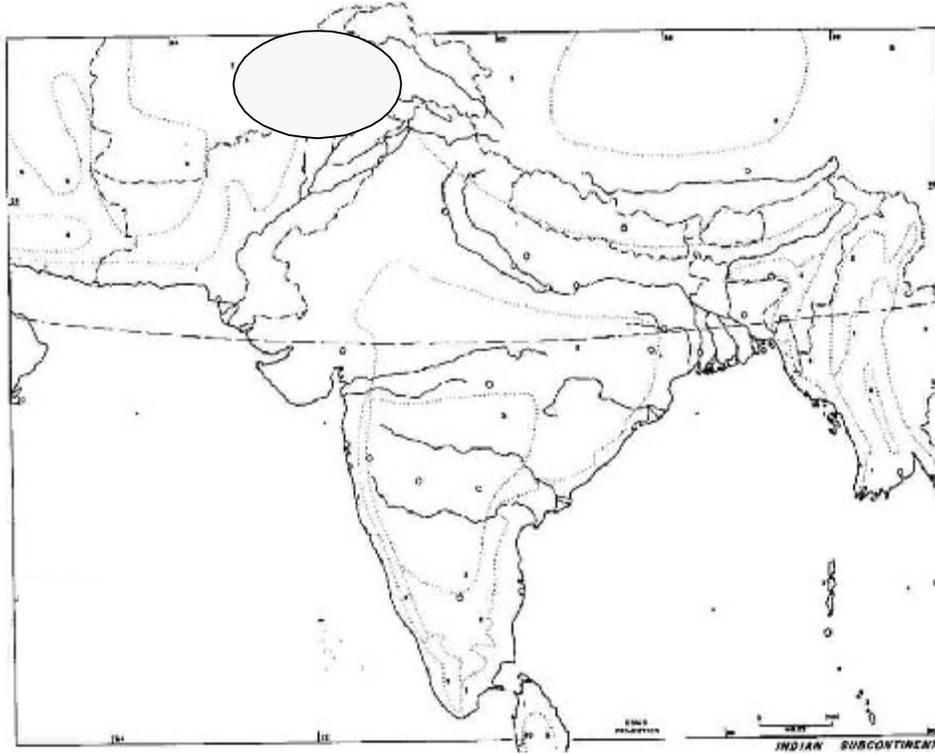
Figure 4: Region with "six" containing *šo-*

7. Numbers of mixed origin

The numbers (*j*)*ekh* "one" and *pandž* "five" cannot be entirely identified with one specific Indo-Aryan subgroup. However, each number bears more similarities with one subgroup than with the others with which it is related.

Ekh and *jekh* are mainly of Central Indian origin. However, the aspiration in these two forms of the number "one" in Romani appears to be found (amongst Indian languages) only in Standard Kashmiri (*akh*, *ok″*), in the Kashtawari dialect (*akh*), and in the Poguli dialect (*yakh*), as well as in Eastern Balochi (*yakh*, *ya*) (Grierson 1968a:3). In Figure 5,

all of these languages (in addition to Hindustani, which uses the more common *ek*) are represented.
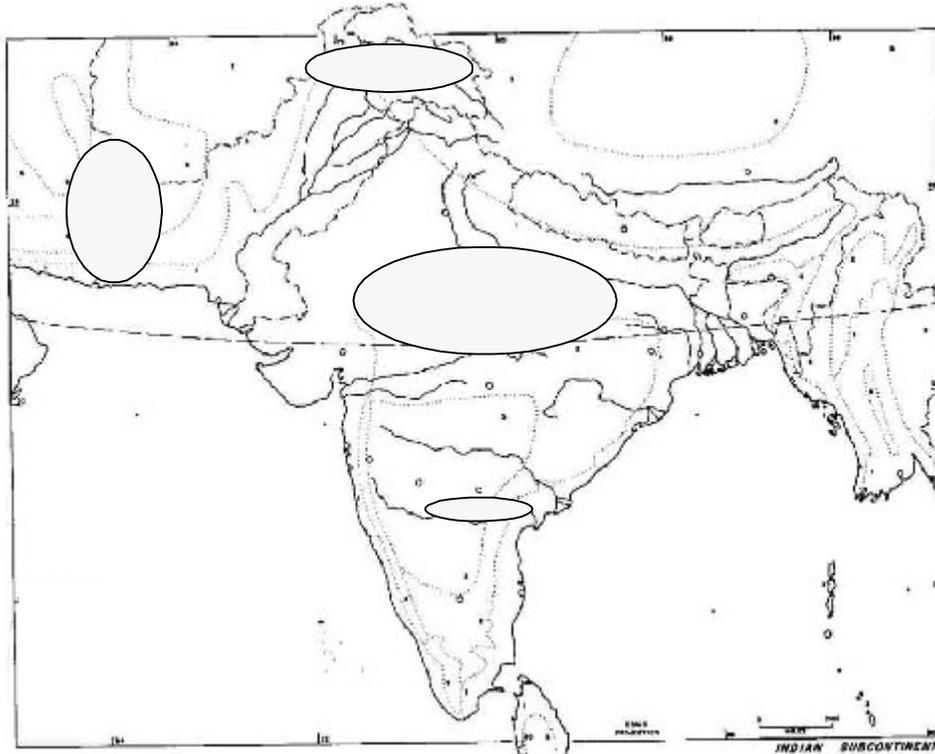

Figure 5: Principle regions with "one" as *(y)ek*, *akh*, *ok''*, or *yakh*

According to Masica (1991:204), there are several other cases in which voiceless unaspirates in MIA are converted to "aspirates in the *final* position" in Kashmiri. In fact, "…all unaspirated voiceless stops are replaced by aspirates in this position…". Considering also that the word *jek* also exists in Romani, the Proto-Romani word for one was probably *(j)ek(h)*, i.e. with the basic form [ek], but also with [jek] or (once Dardic languages began to influence Romani) [ekʰ] and [jekʰ] as variations.

Though *pandž* is most similar to the words for "five" in Iranian languages, it is also quite

similar to its Indic equivalents. In fact, the words for "five" in Punjabi, Gujuri of Hazara

(a Rajasthani language), some Pahari languages[5], and the Garwi and Torwali languages

of Kohistan (almost) exactly resemble the Romani equivalent (Grierson 1969a:11).

Therefore, it, too, is most probably of Central Indian origin. Perhaps an Iranian-based

sound change (from Sanskrit *pañcha* to Persian/Farsi *panǰ*) helped to make it the most

established word for "five" in Romani. In figure 6, Persian/Farsi and the Indian languages

with *panǰ* as five are designated by region.
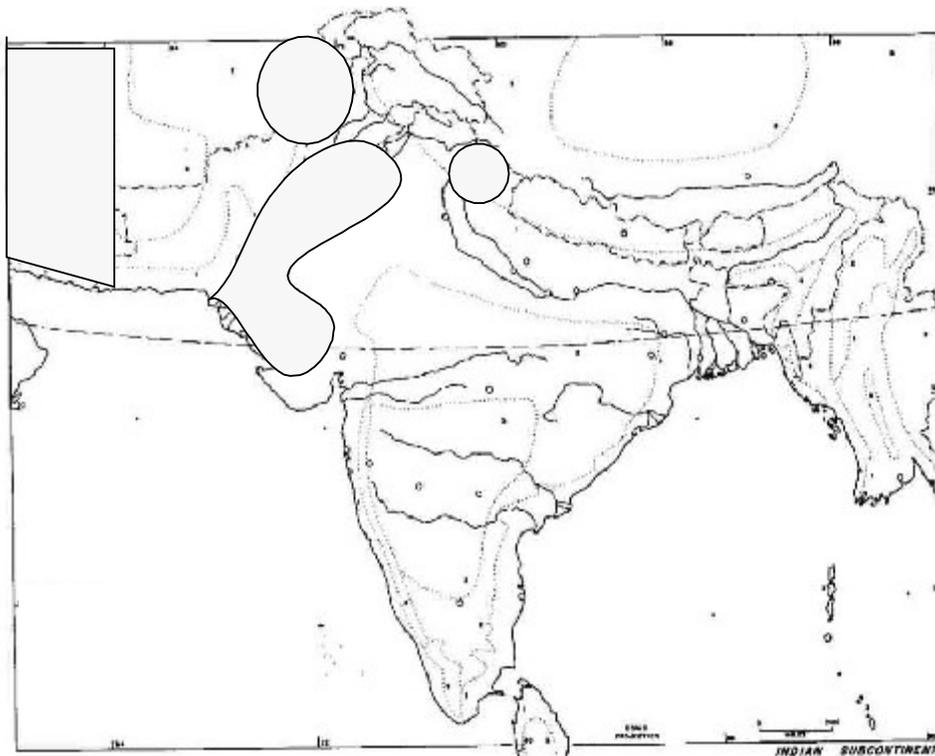

Figure 6: Regions with "five" as *panǰ*

The "voicing of voiceless stops after nasals" (Masica 1991:203) that occurs in *pandž* was

almost certainly introduced into Romani before the ancestors of the Romani people left

India. The Proto-Romani form, therefore, was probably identical to the current form (in some dialects of Romani), i.e. *[pãndʒ].[6] It is also worth noting that most of the above-mentioned Indian languages are spoken by Punjabis and Rajputs—the two ethnicities that are, according to Vijender Bhalla, most closely related to the Romanies (Bhalla in Hancock, *We* 9).

The words for "one" (even with the aspirate *kh*) and "five" are mainly of Central Indian origin. However, it seems likely that both of these might have changed in pronunciation due to influences from other subgroups of Indo-Iranian languages: Dardic in the case of *(j)ekh* and Persian/Farsi in the case of *pandž*.

8. Numbers of questionable and highly mixed origin

The three numbers that have not yet been discussed, namely *deš* (ten), *biš* (twenty), and *šel* (hundred), could come from a variety of sources. Based on the equivalents of these numbers in modern-day Indo-Iranian languages, the origin of these numbers is unresolved and quite questionable. The first two of these three are partly Iranian and partly Indic (where "Indic" refers to either the Central or the Dardic subgroup). *Šel* could come from Sanskrit, Dardic, or both—and also includes a similar Iranian influence to that found in the word *deš*.

No Indo-Aryan or Dardic languages include /e/ in their words for ten. However, in Persian/Farsi, there is no short vowel [a]—only [æ], which Indians could easily confuse

with [ɛ] due to similarity in sound (and because Hindustani, for example, may include [ɛ] but not [æ]). If we thus replace /e/ with the /a/ that is more commonly found in Indian languages, we find many languages that have the word *daš* mean "ten" (see figure 7): Dardic languages (Kalasha, Gawar-bati, the Kohistani languages) and Western Pahari languages (Sirmauri, Baghati, Kiuthali, Pangwali, and Bhadrawahi [Grierson 1969a:21]).
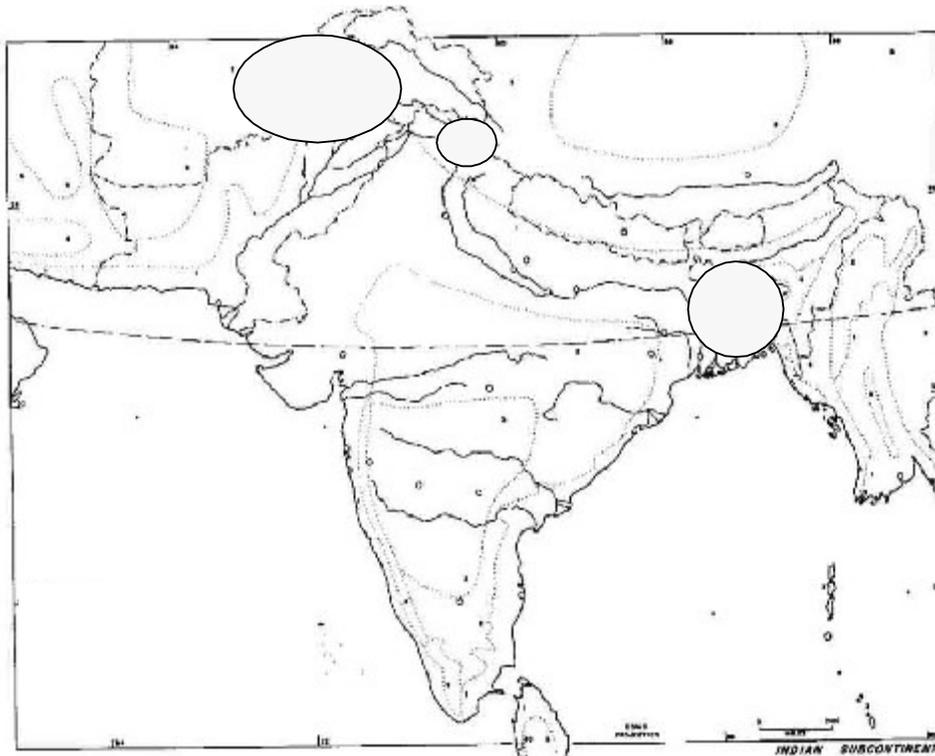

Figure 7: Regions with "ten" as *daš* and "twenty" as *biš/beeš*

In Romani, Sanskrit *ś* (as in *daśa* "ten") often changes to *š*; this happens in many Indian languages as well, though different languages implement this change with different words. In Hindustani, for example, Sanskrit-derived words like *śrī* (an honorable title

meaning something like "lord") are pronounced e.g. *šrī* [ʃriː]. Though there are different

letters for *śa* and *ša* in the Devanagari script for Hindi, both are pronounced [ʃa] in

practice[7]. The same applies to Bengali, in which the word is spelled *daś* based on Sanskrit

but is pronounced [ɖɔʃ].

The situation with the word for "twenty," *biš*, is similar in terms of linguistic regions in

which cognates are used (see figure 7). In contrast to *deš*, it does not include any sounds

that are missing in Indian equivalents: the Sanskrit word *vimśatih*, once again, includes *ś*.

This has been converted into the Kohistani languages and some Pahari languages

(Jaunsari, Shodochi; and once again, Sirmauri, Baghati, and Kiuthali) as the *š* that is

found in Romani. The word for "twenty" in these languages is *bīš*, i.e. apparently

pronounced [biːʃ] (Grierson 1969a:23). Again, many dialects of Bengali also use the word

biš, but a Dardic or Northern Indian origin is more probable than an Eastern Indian origin

considering the main area inhabited by the Romanies' ancestors (which was not quite as

far east as Bengal).

The word for "hundred," *šel*, is consistent with the sound changes from Sanskrit *ś* to

Romani *š* and from *t* to *l*: the original word in Sanskrit was *śatam*. On the other hand, it is

also very similar to its equivalents in some of the Shina languages of western Kashmir

(*šal* in Gilgiti, *šall* in Chilasi, and *šāl* in the Dras dialect) and in Maiyã (*šal*), all outlined

in figure 8 (Grierson 1969a:27). It may even be possible that the Romani word has its

origins in both; it may have been introduced into the language by Sanskrit and then

reinforced by some Dardic languages. This word also includes the Iranian-based sound

change from Indic [ə] to Persian-based [ɛ] found in *deš*.
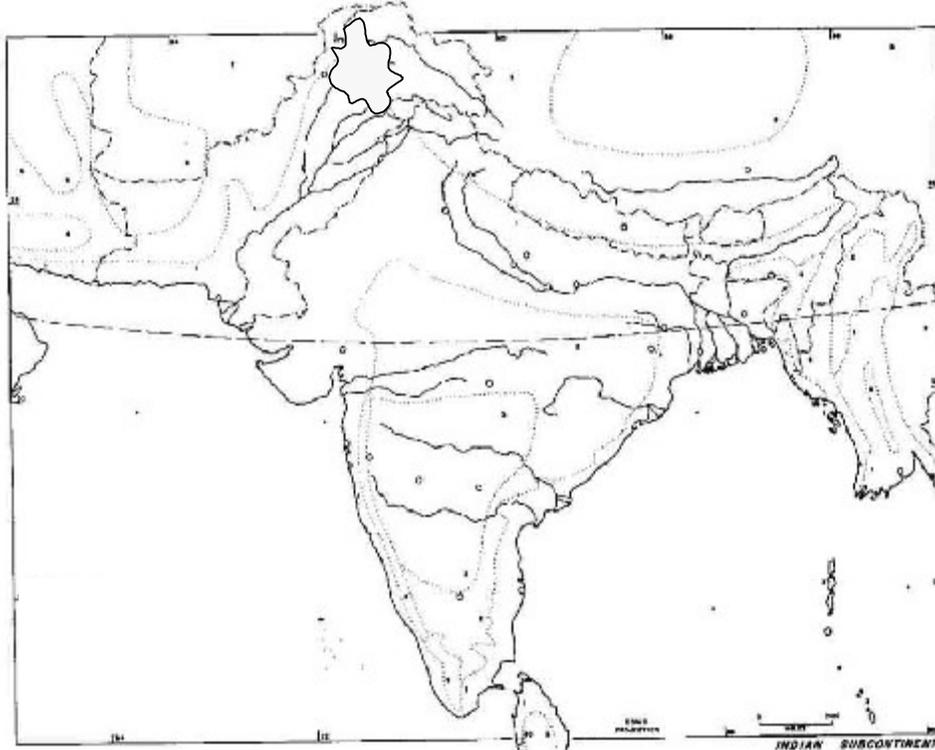
Figure 8: Region with "hundred" as *šal*

Masica includes a more detailed study of the indirect sound change from Sanskrit *śatam*

to the Dardic (and Romani) forms including -*l*- in lieu of -*t*-; the excerpt provided below

includes only a small amount of additional information:

> Intervocalic -t-, -d-, lost in the Literary Prakrits except for Sauraseni [Central Indo-Aryan
>
> languages], lasted longer in the northwest; -t- and possibly also -d- are represented by l
>
> or r in modern Romany dialects…and -t- only, by r in Khowar and l (sometimes evolving
>
> further into u) in Kalasha…OIA *śata*…> Kal[asha] *ʃau* (<[*]*ʃal*, found as a loanword in
>
> Shina)…[Masica 1991:199].

Many modern Central and Mediate languages, however, include very different forms derived from the original Sanskrit (e.g. *sau* [sɔ], sometimes [saw], in Hindustani). Hindustani (as well as some other languages of this region, perhaps) was only being formed from MIA when the Romanies' ancestors left India. It is probable that this exodus (*Teljaripe* in Vlax Romani) had occurred before a sound change from *ś* to *s* became established in the regions of Central and Mediate languages. Thus, the Proto-Romani form may have been **šat(a)*, until Dardic influence converted the *-t-* to *-l-* to create **šal* —or perhaps **šal* was simply a loanword from Dardic that was later influenced by Iranian as well.

There are some Romani numbers that can definitely be identified with one linguistic subgroup. However, three out of the nine numbers of Indian origin could be associated with multiple subgroups. *Deš* could be of Dardic or Western Pahari origin; unlike *biš* and like *šel*, it also seems to include a Persian-based sound change from [a] to [ε]. *Šel* could originate in Sanskrit, the Dardic languages, or both. All three numbers are very ambiguous as far as linguistic origin is concerned.

9. Origins for Romani synonyms

In addition to the numbers, many synonyms of different origins exist in Romani. Each synonym comes from a different Sanskrit root. Some of these synonyms have a clear source in Romani; others are less certain and may be of mixed origin. Some of the words listed below have no clear etymology whatsoever.

The origins of and cognates for each of these words was found in Turner's dictionary (1969:131-760):

"to wash" (verb-stems): *thov-*, *xalav-*

"to burn": *thab-*, *thar-*, *phab-*, *xačar-*

"bracelet": *kero*, *merenklo/mereklo/merikli*

"to fear": *dar-*, *traš-*

"liver": *buko*, *kalindžo*

"to move": *čalav-/čarav-*, *kar-*

"to open": *puter-*, *p(h)(i)rav-*

"to sing": *bag-*, *gab-*, *gilaba-*, *kixiv-*

"tomorrow": *kaliko*, *tehara*

"wet": *kingo*, *sapano*, *tindo*

The verb stem *thov-* can be found under the Sanskrit entry *dhauvati* "washes" and has cognates in a wide variety of Indic languages, Dardic languages as well as Indo-Aryan languages. Most cognates share the Romani meaning "wash"; some others have related meanings, e.g. Bihari *dhoeb* is glossed as "to beat (clothes in washing)." By contrast, the verb stem *xalav-* is listed under the entry *kṣāláyati*, which also means "washes." It, too, has several cognates in modern Indian languages[8]; however, most of its cognates mean not "to wash" but "to rinse." The cognate most closely related to Romani *xalav-* is Garhwali *cāḷnu* "to wash." However, Turner does not list any Garhwali descendants of *dhauvati*, so the one language that has a close cognate of Romani *xalav-* appears not to have a cognate of its synonym *thov-*.

Both *thab-* and *thar-*, meaning "to burn," are listed under *dagdhá-*. However, Turner

explains that *thab-* is specifically derived from MIA **dabbhaï* "after type *laddha-:*

labbhaï" (1969:349). The only cognates provided after this derivation are cognates in

other dialects of Romani and the masculine Gujarati noun ḍ*ābh* "brand, cautery." Thus,

this Gujarati noun may be the closest cognate to *thab-* extant in modern Indian languages,

though more cognates may well have existed at the time of the Romani exodus or

Teljaripe.

Phab- is instead listed under *bhabh*. Turner originally listed only Vlax Romani cognates

in this entry; however, a newer, online version instead lists a few Indic cognates, as

follows:

> S. *bhabhaṛu*, *bhabhaṭu* m. ' flare '; N. *bhabāunu* ' to burn (esp. with pain) '; -- ext. -- *kk* -- : S.
>
> *bhabhaka* f. ' flare '; P. *bhabak* f. ' glare '; H. *bhabaknā* ' to burst into flame '; G. *bhabhkɔ* m. '
>
> threat '; M. *bhabakṇẽ* ' to rise (of fire) ' (2006:534).

Evidently, although *thab-* and *thar-* have (like *thov-* and, in some sense, *xalav-*) cognates

in a wide variety of Indian languages, including Dardic languages (e.g. Kashmiri *dodu*

"burnt"), *phab-* must have an Indo-Aryan source, since it has no Dardic cognates.

However, outside of Turner's dictionary, we may also find the etymology of the verb

stem *xačar-* (also meaning "to burn"). It is from Old Indo-Aryan (OIA) **[ṭṣaːṭi], meaning

"burning" (Hancock 2006:28).[9] The cognates of the Romani verb are mainly in Indo-

Aryan languages; however, like *thar-*, *xačar-* also has cognates in Shina and Kashmiri as

well. Even if Gujarati has cognates for all of these words (as suggested by the newer edition of Turner's dictionary), the cognate for *phab-* (listed above) means "threat," not "to burn," so it must be at least somewhat different from Romani *thab-*, *thar-*, *phab-*, and *xačar-*.

Kero "bracelet" is listed under *káṭa(-)* (to be distinguished from the next entry *kaṭa*, which has no stress on the first syllable). In Sanskrit, similar words (or this word itself, possibly in conjunction with similar words) were used in various sources to mean "twist of straw," "mat," "bridle ring," and "straw mat," as well as "bracelet," which, as Turner points out, is the only meaning retained in NIA:

> In NIA, except in G[ujarati], *káṭa-* has developed as 'ring, bracelet, chain' [Turner 1969:131].

This word, too, has equivalents in both Indo-Aryan and Dardic languages. However, note that it, too, observes the *a > e* sound change found in *deš* and *šel*.

The etymology of *merenklo* is not too clear. Turner lists it under *maṇí(-)* (again, to be distinguished from *maṇi*), which means "jewel" or "ornament" in Sanskrit as used in the Rig Veda. The cognates listed usually have meanings such as "pearl," "coral," and "bead" rather than "bracelet." However, the Greek Romani word *minrikló* is glossed as "ornament," which is also one of the glosses for *maṇí*, not "bracelet." This word probably evolved in meaning within Romani itself, but it is still unclear how *maṇí* could have phonologically yielded *minrikló*. Perhaps it evolved from the related word *māˊṇikya* "ruby"; perhaps the ending *-ya* was viewed as an adjectival suffix corresponding to

Romani -*lo*. In any case, the precise etymology remains a mystery, but most probably, no other Indian language has two words for "bracelet" with one derived from *maṇi*.

According to Turner (in his article "*Traš*-, 'To Frighten'"), Miklosich traces the verb *traš*-, which means "to fear" in Vlax Romani, to Sanskrit *trásati* "to tremble." Turner, on the other hand, remarks that Miklosich's derivation fails to explain Romani *š* and instead traces it to an earlier (OIA?) form **tarṣáyati*, which evolved by anticipation into a verb stem **trarṣ*- and then **trašš*-. He notes:

> In Kāśmīrī, Lahndā, and Sindhī, in words which begin with a dental and in which the first syllable ended with *r* or the second (and possibly the third) syllable began with a consonant group containing *r*, by a process of anticipation the *r* was pronounced also immediately after the initial dental [Turner 1924:39].

In the *Comparative Dictionary of the Indo-Aryan Languages*, under the entry *trásati*, we see that the cognates of Romani *traš*- are no exception. Many of them, especially those in Dardic langauges, also begin with the cluster *tr*- (e.g. Kashmiri *trasun*- "to be afraid," but also Gujarati *trāsvũ*, a [neuter?] noun meaning "fear").

Traš- has cognates mainly in Dardic languages but also in a few Indo-Aryan languages (e.g. Lahnda, possibly Sindhi, as well as aforementioned Gujarati). Its synonym, *dar*- ("to fear"), has many indirect cognates (somewhat like *xalav*- "to wash"), but Turner lists only Gujarati as having any direct cognates with the Romani word. Turner lists Vlax Romani *dar*- (as well as *darav*- "to frighten") under Sanskrit *dārayati* "tears asunder, scatters"; its Gujarati cognates are ḍ*ārvũ* "to frighten, threaten, stop" and ḍ*ārɔ* "threat (masc.)." Just as many other (apparently less directly related) cognates of *xalav*- "to

wash" are listed under *khakkhālayati* "rinses," so many other such cognates of *dar-* are listed under *darati*, which indeed originally meant "is afraid" (much closer to the modern meaning of *dar-*, and of its [also less directly related?] cognates, than "tears asunder.") As in the case of "to burn," Gujarati has cognates for both *dar-* and *traš-*, but not as synonyms per se.

Turner (2006:698) lists the word *buko* under the entry *vṛkká* "kidneys" but actually traces it to a related word *bukka*, meaning "heart," though another gloss provided for this word (apparently based on the speech of a much earlier period) is more general: "intestines (heart, liver, lungs, &c.)." The meaning "intestines" is preserved in Greek Romani *bukó* and Sindhi *boku*. However, Romani is the only language in which cognates of *bukka* or *vṛkká* are glossed as "liver (in all dialects listed except Greek Romani)," "spleen (German Romani *pukko*)," or "lungs (*pukko*, Romanian [Vlax?] Romani *parno buko*—literally "white *buko*," as opposed to *kalo buko* "black *buko*" or "liver"—and the plural form of Spanish Romani *buko*)," though almost all Indo-Aryan languages listed (except Sindhi) have a cognate meaning "breast" or "chest." However, *pukko* in German Romani also means "kidney" (as well as "liver," "spleen," and "lungs"), and *buko* in Spanish Romani means "courage" as well as "liver." Lahnda, Kumauni (a Pahari language), Kashmiri, and Torwali (another Dardic language) each have a cognate meaning "kidney(s)," and the meaning "courage" is shared in Bengali and Oriya.

By contrast, almost[10] all of the cognates of Romani *kalindžo* are in Indo-Aryan languages (though Kashmiri borrowed the Lahnda cognate) and mean "liver," though some

cognates have other meanings as well (e.g. Hindustani *kalejā* means "human liver" but also "vitals" or "courage"), and some languages that have cognates meaning "liver" have other cognates that do not (e.g. Hindustani also has *karejī*, meaning "meat round the vitals of animals considered good for eating").

Vlax Romani *čalav-* and Welsh Romani *čarav-*, both meaning "to move, stir," are listed under *cāráyati* "causes to move, shakes, grazes," along with *čar-*, an intransitive verb meaning "to graze, eat" in Romani. The transitive Vlax Romani verb *čalav-*, meaning "to move, touch, strike," is listed under *cālayati*. Both have cognates in a variety of Indian languages (though e.g. Bengali and Oriya cognates are listed only under *cālayati*). Different cognates in Hindustani can be found in both entries: under *cāráyati*, a cognate in Old Hindustani, meaning "to graze cows," is listed (as well as a word in modern Hindustani formed as the causative of the old Hindustani verb); under *cālayati*, a cognate in modern Hindustani is listed meaning "to shake, sift." Words such as Hindustani *calnā* "to move, go on or away, be in use," which may appear to be cognates, are instead listed under *cálati*, meaning either "goes away" or "shakes, trembles."

Kar- "to move" is listed under *kaḍḍhati* "pulls, draws" and has cognates in a variety of Indian languages, including Hindustani, Gujarati, Assamese, Kashmiri, and Gawar-bati. The cognates tend, however, to maintain meanings more similar to "pull" or "draw" rather than "move." Thus, no cognates have been found that have cognates of both *čalav-* and *kar-* meaning "to move."

The etymology of *puter-* "to open" is unknown, though Sanskrit *sphuṭati* has been suggested to be related. If this is, in fact, correct, it is still unknown how the initial *sph-* was replaced by *p-*, how ṭ was converted to Romani *t* rather than *r* (as in e.g. *káṭa/kero*), and why *-ti* was converted to *-r* rather than *-l* (as in e.g. **tarṣáyati/trašel*).

The case is somewhat similar with the etymology of its synonym, *p(h)(i)rav-* ("to open"). Sanskrit *bhindati*, which Turner glosses as "splits," has been suggested to be related. Turner does not list any Romani cognates under the entry *bhindati*, but he does include some cognates in Indo-Aryan languages.

The word *bag-* "to sing" has many cognates, but only a few of these are derived from MIA **vagga* according to Turner, and these seem to be its closest cognates. These cognates are exclusively Indo-Aryan and include Sindhi *vagaṇu* , Lahnda *vaggaṇ*, Gujarati *vāgvũ* and all meaning "to sound (e.g. an instrument)" (as do most of their less closely related cognates in a wider variety, e.g. Hindustani *bājnā* ["whence *bajnā* ' id. '"] and Kumauni *bājṇo*). Turner also lists cognates in older Indo-Aryan languages: in Old Marwari, *bājaï* meant "to sound," but in Old Awadhi (spoken in Awadh/"Oudh" in Uttar Pradesh), the same word is glossed as "sounds, thunders, roars." Romani is the only language with a cognate meaning "to sing" (and another cognate meaning "song").

German Romani *gab-* is listed under *gāpayati*, and presumably also Vlax Romani *gilaba-* has a similar origin, though it is not listed in this entry of the dictionary. Cognates exist in a wide variety of Indian languages, but in this case, they also share the meaning "to sing"

(with only one very minor exception: in Assamese, in addition to *gāiba* ["to sing," or perhaps "I will sing"], there is another cognate *gāuni* "wages for singing"). Turner also lists Old Marwari *gāve* and Old Marathi, Old Awadhi, and Old Gujarati *gāvaï*, all meaning "sings"; *gāvaï*, in fact, is also shared in Prakrit. These are probable clues to the origin of Romani *gab-*.

In a 19th-century dictionary, Sztojka (1886) lists *kixiv-* as a word meaning "to sing" in Hungarian Romani. This has not yet been found in any other dialects of Romani. However, it has been proposed that it is related to the onomatopoeic Sanskrit verb *kīkkati* "screams," which raises another problem of deriving *-xiv-* from *-kkati* or something similar. Essentially, what we can tell of the origins of the Romani words for "sing" indicates that the Indian languages include neither a cognate of *bag-* that means "to sing" nor any close cognates of *kixiv-* (which we might trace to Rajputic **kikhiv-*).

There are two words for "tomorrow" in Romani: *kaliko* and *tehara*. Turner provides an etymology for neither of these words, though he does provide an etymology for Russian Romani *kal*, which is related to *kaliko*. *Kal* is listed under *kālá-* meaning "time"; *kaliko* might be derived from the related word *kālya-* "timely, seasonable." It has been suggested that *tehara* is related to Sanskrit *divasa*, but again it is difficult to account for the sound changes in this case.

There are also two words for "wet": *kingo* and *sapano*. Turner does not suggest an origin for either of these words, and *sapano* is of uncertain origin. He suggests an origin only

for another word *tindo*, which he lists under *timita-* as the Vlax Romani word meaning "wet." *Kingo* and *tindo* appear to be derived from the same root and may be dialectical variations. Turner also lists *čindo* in Hungarian Romani and *kindo* in German Romani, Russian Romani, English Romani, and Welsh Romani. This kind of variation is common between Romani dialects; a similar example is the word for "your (masc.)," which is *tiro* for some Romani-speakers, *tjiro* for others, and *kiro* for still others. Perhaps *kingo*, in turn, is a variation of *kindo*.

For *sapano*, Tálos (1999:251) suggests an improbable origin by first noting the formation of the Sanskrit adjective *sārdra* "wet, moist" from the comitative prefix *sa-* ("with") and the word *ārdra*, which also means "wet." He then suggests that *sapano* was formed for a combination of Sanskrit *sārdra* "wet, moist" and *āpya* "living in or belonging to water." I find that a more probable origin may be found in the Sanskrit word *sapānīya* "wet" (cf. *sa-* + *pānīya* "water"). Though Turner lists only two Dardic cognates for this word, neither of which sound similar to the Romani *sapano*, this may be a word that disappeared from the vernacular languages of India. Some other words with the same prefix survive in Indian languages, e.g. words beginning with *saphal-* meaning "success" (cf. Sanskrit *saphala* > *sa-* + *phala* "fruit").

In summary, though the origins of some of the Romani synonyms are unclear, no group of synonyms has been convincingly traced to any one Indian language. None of these groups of synonyms exists in any other Indian languages; to match any group to an Indic

language necessarily entails, at the very least, semantic changes between Romani and the Indic languages that have not been explained.

10. Lomavren

The language of the Lom people is essentially Armenian with some words of Indic origin. It may have originated as a koïné, as Romani did. Unlike Romani, though, Lomavren seems to be distinctively Indo-Aryan; where Romani may have Dardic words, Lomavren tends to instead use an Indo-Aryan equivalent. Often words in Lomavren are closer to their equivalents in e.g. modern Hindustani than Romani words.

Two words of Dardic origin present in Romani, but not in Lomavren, are *phabaj* "apple" and the number for "hundred," *šel*. The origin of *šel* has already been discussed above. *Phabaj* is related to the Sanskrit equivalent *bhabbā*, which has cognates almost exclusively in Dardic languages (a cognate might exist in Domari) and no cognates in modern Indo-Aryan languages spoken in India. But the Lomavren equivalent is *ansev*, similar to Hindustani [seʋ] which is a loanword from Persian (the word for "apple" in Modern Persian is [si:b]). It is possible that the Lomavren word was formed via metanalysis; [ãn si:b] in Persian means "that apple," but it (or its archaic form) may have been mistaken to be one word, hence *ansev* rather than simply *sev*.

The Romani word *but(j)i* might be another example. It is from Sanskrit *vrtti* "mode of life, conduct, business, wages." Though it has cognates in Indo-Aryan languages as well (e.g. Hindustani *buttī* "means of subsistence" and *bīṭ* or *bīt* "grazing fee charged by herdsmen"), the language that seems to have the closest equivalent to Romani *but(j)i* is Kumauni, in which both *but* and *buti* mean "domestic work." Lomavren instead has *kam*,

and the word for "work" in Hindustani is also [kã:m]; both can be traced to Sanskrit *kárman* "act, work" and have a wide variety of cognates.

There are other words in Lomavren that are more similar to their modern Indo-Aryan equivalents than the Romani equivalents. For example, *dad* is the Romani word for "father" and is of Indic origin, but Lomavren *bap* is nearly identical with modern Hindustani [ba:p]. *Dad* comes from Sanskrit *dādda* "father or other elderly relative." The only languages that preserve the specific meaning "father" are three Dardic/Nuristani languages: Ḍumaki, Dameli, and the Gulbahari dialect of Pashai. Lomavren *bap*, on the other hand, is derived from *bāppa* meaning "father," and many modern Indian languages have cognates with the same meaning.

A more extreme example is the words for "foot." In Romani, it is *pi(n)rro*, but in Lomavren it is *pav*, and in Hindustani it is [pã:ʋ]. *Pinrro* is from Sanskrit *píṇḍa*, which variously meant "lump," "calf of leg," or sometimes simply "piece." No cognates have been found with the precise meaning of "foot," although some mean "leg," and the Churahi dialect of West Pahari has both *phinī* "ankle" and *pindā* "leg."

Perhaps the most extreme example found by far is "grease." The word for "grease" in Romani is *čiken*, but it is related to Sanskrit *cikkáṇa*, which is not even a noun but rather an adjective meaning "slippery," "unctuous," or "stuck to." Not only is Romani the only language to have a cognate meaning "grease," but in fact only two other languages (the Eastern or *Māgādhi* languages Oriya and Bihari) even have cognates that are nouns

(meaning "oil [in Oriya]," "oil seeds in general [in Bihari]," or "ghee [in both Oriya and the dialect of Bihari spoken in the city of Gaya]"); all other cognates are adjectives, like the original Sanskrit. In Lomavren, the word for "grease" is *tel* (cf. *tailá* "[sesame] oil"). In fact, Turner (2006:340) glosses the Lomavren word as "oil, fat, butter," and indeed, it has a variety of cognates, almost all of which mean "oil" (the others mean something similar, e.g. Bengali *tēlā* "oily." Bengali also has *tel* to mean "oil").

There are also situations of the opposite type, in which Romani has a word that could be of Indo-Aryan origin and Lomavren has either a word of Dardic origin or a relatively uncommon word. One example is the word "dog," which is *džukel* in Romani. Lomavren has another word, thought to be of Indic origin: *solav*. Turner (1969:739) suggests that it may be from Sanskrit *śvāpad*, but he admits that this is "very doubtful" and provides no other cognates. Bakker (1997:9) gives a more convincing argument for a Dardic origin: modern Lomavren *solav* may have originally been **sodav* (d > l is a common phonological change in Armenian and, by extension, Lomavren as well), and there are many Dardic languages with similar words, e.g. Kalasha [ʃuːɽæ].

By contrast, *džukel* is from OIA *jakuṭa*; many cognates have been found in some Indian languages that are now relatively obscure, apparently Indo-Aryan "Gypsy" languages spoken in India. The most similar cognate is Myaanwaale *jukēlā*, but other cognates include Garodi *dẕā*, Kanjari (Sitapur dialect) *jhūm̐kil*, and Nati (Rampur dialect) *jhumkar*.

Another example is the word "bird," which is *čhiriklo* in Romani and *pantri* in Lomavren. According to Turner (1969:248), *čhiriklo* is from Sanskrit *caṭaka* "sparrow," as is Hindustani [tʃiɽija] "bird." Cognates exist in a variety of languages, even Burushaski. But Turner (1969:437) classifies Lomavren *pantri* under the Sanskrit word *pattrin*, meaning "leaflike, thin" as well as "winged, feathered" and "bird." Lomavren is the only language listed as having the meaning "bird"; the other cognates are more closely related to the meaning of "leaflike, thin" (they usually mean "thin," but *pati* in Assamese, Bengali, and Oriya is instead glossed to mean "a little"). The latter cognates are somewhat related to Romani *patrin* "leaf," but Turner lists this Romani word under *páttra* "wing(-feather), leaf, leaf of book, petal."

There are two words for which Indo-Aryan cognates can be found in both Romani and Lomavren, but in these cases, the Lomavren word seems to be an extension of a meaning that might be described as more general. The first word is the word "village": *gav* in Romani and either *lei* or *les* in Lomavren (*les* may also or actually mean "world" in Lomavren). *Gav* is from Sanskrit *grāˊ ma* "village" and has a wide variety of cognates. The Lomavren word is derived from another Sanskrit word *dēśá* "point, region, part, province, country"; some Indo-Aryan and Dardic languages (including Hindustani but not Kashmiri) have cognates, e.g. Hindustani *des(h)* "country," and in some Indian languages, the meaning of this word is extended to mean "native village" or "native country" as well.

The second word is a more definite case. The verb stem meaning "to buy" in Romani is *kin-*, which is derived from Sanskrit *krīṇā́ti* "buys." Its Indo-Aryan cognates usually mean "to buy"; its Dardic cognates usually mean the opposite, "to sell." The word *krenik* in Khowar (another Dardic language) also means "to buy." The Indo-Aryan languages are usually Eastern languages with the exception of Hindustani, which also has a cognate *kīnnā* "to buy (stem *kīn-*)."

In Lomavren, however, the verb stem meaning "to buy" is *li-*, which also means "to take." It comes from Sanskrit *lábhatē* "catches, takes," and even has a cognate in Romani (as well as many others in a variety of Indic languages): the verb stem *le-* in Romani also means "to take." The original meaning of the Lomavren word was probably "to take," but the meaning was later extended to also mean "to buy."

Finally, there are some words in Romani and Lomavren that do not fit any of the above-listed categories. One example is the words for "middle": *maškar* in Romani, *mandž* (or *manj*) in Lomavren. Both words have the same origin and can be traced back to their Sanskrit equivalent, *maddhya*. In fact, the word for "vulva" in Romani is *mindž*; this may have formed either as a euphemism (if it came from *maddhya*) or else could be a lexical borrowing from the Lezgian equivalent *miš* (Hancock 2006:15).

The word for "fear" in Lomavren may be a more interesting case. The etymology of Romani *dar-* has already been discussed, but Lomavren instead has another verb meaning "to fear": *bi-*, which is from Sanskrit *bhīyatē* "fears." All of its cognates are in Dardic

languages that are now relatively obscure, e.g. Phalura. However, *bhīyatē* is related to the word *bhayá*, which means "fear" in Sanskrit but is found even in unrelated Indian languages (e.g. Dravidian languages) as a loanword. Thus, though its direct cognates may be strictly Dardic, it is also related in some way to words in a much wider variety of words than is the Romani verb stem *dar-*.

11. Further work

This paper is a work in progress; more detailed and more thorough investigation is required in order to provide an elaborate study of the Romani numbers and their original forms. Some of the questions that remain to be solved concern details in lesser known languages of the Indian Subcontinent; others are much larger, referring to an explanation of what has caused the diversity of Romani origins.

Whether or not there are any Indian languages that use *dui* and *do* depending on grammatical function is a question that requires further investigation. Certainly, Hindustani does not do this; a noun may change case but never an adjective.

The origins of some of the NIA cognates of Romani words need to be investigated further. For example, in the case of the number "four" (*štar* in modern Vlax Romani), there could be some dialect of Wai-alaa that includes the *r*. Perhaps an earlier or archaic form had a final *r* that was dropped.

Another phonological phenomenon has been mentioned elsewhere: there are many words in Romani with cognate forms in Sanskrit such that the Romani form replaces Sanskrit [ṭ] with [ɫ]. In the case of the number *šel,* some Dardic languages also have [l] rather than the original dental. It is not yet known whether or not there are any other words in Dardic languages that, like their closest Romani equivalents, have undergone this sound change.

Rajputic forms have been suggested for only a few numbers at this point. More numbers will be found, but more information concerning sound changes from MIA to Romani is needed.

No explanation has yet been provided for the fact that the numerals "seven," "eight," and "nine" are not of Indic origin in Domari, Lomavren, or Romani. In languages such as Hindustani, which have fusional rather than analytical forms of numbers, two-digit numbers ending in "7" or "8" tend to follow certain patterns (which do not seem to be followed by other numbers, especially numbers with the same initial digit). Also, almost all two-digit numbers ending in "9" follow the pattern "one taken away from multiple of ten," with the sole exception of 99 [nĩnã:ʋe]. This requires further and more thorough investigation and should probably be written in a different paper entirely.

The ancestors of the Romanies left India around 1000 A.D., as the Hindustani language was forming. It is possible that Romani can be compared to other contact languages originating in, but fully formed outside, a region in which another contact language was being formed. This idea requires an in-depth analysis into other contact languages.

Perhaps, contact languages like (but other than) Romani have dealt with their numbers similarly; they may have influences not reflected in the "abandoned" language that was already being formed (Hindustani, in the case of Romani).

12. Conclusion

Some numbers are clearly from one group of Indian languages; a few others can be most closely, but not entirely, traced to another group; and still others cannot be identified with any one group. There is no one language that includes numbers similar to all of the Indic numbers in Romani. In fact, there is no one subgroup that is directly related to all of these numbers.

The Romani numbers, therefore, indicate that Romani is not the descendant of any single Indian language. The only possible explanation is that it must be a koïné, descended from and influenced by several Indo-Iranian and non-Indo-Iranian languages.

13. Appendix: Languages and Locations

13.1 Koïné languages

The term "Hindustani" is used in this paper to refer to *Khariboli (Bāzārī) Hindustānī*, the standard dialect of Hindi/Urdu that is used and understood in common speech throughout the "Urdu"-speaking regions of Pakistan and "Hindi"-speaking regions of India. In India, Hindustani formed as a result of contact between Muslim (foreign) invaders and Hindu (Indian) subjects. Both Hindustani and Romani formed from a contact language, designated here by the name "Rajputic."

The Vlax dialect of Romani originates in Romania (the word "Vlax" being derived from the name of the southern region of Romania, *Wallachia*). It is the most widely spoken dialect of Romani (and perhaps even more widely understood, considering its similarity to Balkan Romani dialects), and it is thus accepted as a mainstream form of the Romani language.

13.2 Locations

Chitral is a part of northwestern Pakistan on the border with Afghanistan. It is the northernmost district of Pakistan's North-West Frontier Provinces.

Himachal Pradesh is an Indian state. It is south of Jammu and Kashmir, east of Punjab and north of Delhi. The Pahari languages are primarily spoken in Himachal Pradesh.

Kunar is a province in Afghanistan on the border with Pakistan.

Rampur is apparently a city in Uttar Pradesh. However, there is also a town in Himachal Pradesh called Rampur.

Sitapur is a district in Uttar Pradesh.

Uttarakhand is a state directly north of Uttar Pradesh and south of Himachal Pradesh.

13.3 Other languages

Garhwali is a Pahari language spoken in Uttarakhand and Himachal Pradesh.

Gawar-bati is a Dardic language spoken primarily in Kunar.

Kalasha is another Dardic language spoken in Chitral.

Kanjari is the only "Gypsy language" of India that is known to be an Indo-Aryan language.

Marwari is a Rajasthani language, i.e. a Central Indian language spoken in Rajasthan.

14. Endnotes

14.1 Acknowledgments

I wish to thank Prof. Ian Hancock for his enthusiastic support and for the useful comments and modifications he has made on this paper. In addition, he has been kind enough to provide me access to many useful materials, some of which are cited here as references. I would also like to thank Dr. Sadaf Munshi for her feedback, suggestions, and help regarding the establishment of the relationship between Romani and Dardic languages.

14.2 Notes

[1]The Dom and Lom peoples are not to be confused with the Romani people, though they are often popularly identified as "Gypsies."

[2]*Var* is an Indic word which also means "time(s)" in the sense of "occurrence(s)"; it has close cognates e.g. in Hindustani (Hindi/Urdu) *baar*, Gujarati *vaar*, and Punjabi *vaaree*, all of which have the latter meaning of *var*.[4]

[3]To provide a few examples of kings who married Brahmin wives: King "Shalivahana" Gautamiputra Shatakarni of Pratisthana (present-day Paithan) in Maharashtra and the King of Thirumangai in Tamil Nadu.[5]

[6][ʈrĩːɳi] is the neuter form. It also may be used as a name for the number when counting, though the evidence for this is sketchy. Unlike Hindustani, Romani has other words for

names of numbers, e.g. *trinali* "the number three." However, another characteristic that is shared between Romani and Hindustani is a set of special words for such phrases as "both," "the or all three," "the or all four," etc. In both languages, these words are formed by combining a number with a plural marker (in Romani, the plural marker is the prefix *li-*; in Hindustani, it is the oblique plural suffix [õ]), e.g. Romani *li-trin* = Hindustani [ṭĩ:nõ] = "the three, all three."

[8]Specifically: Kiunthali (which also has *pānz* for "five"), Shodochi, Mandeali, Chameali, Gadi, and Pangwali.

[9]Many dialects replaced [dʒ] with [z], so *pandž* is often pronounced [pãnz] instead of [pãndʒ].

[10]Compare this to South Indian languages (e.g. Malayalam and Telugu), in which *ś* is pronounced /ɕ/ and contrasted with *š* /ʃ/.

[11]All but four of these cognates are listed under a related Sanskrit verb **khakkhālayati*; two of the remaining four are the Prakrit forms given above. The third cognate is in the Bhalesi dialect of West Pahari: *chāli deni* "to wash clothes by treading them." The final cognate is in Kumauni, another Pahari language: *khālṇo* "to empty out (water)."

[13]Turner cites this word as being present in the Rig Veda with this meaning, under entry #5364 *jhāta* "burnt."

[14]The one exception is Assamese *kɔlizā*, which means "heart." The meanings of "liver" and "heart" are often confused in Indian languages (e.g. in Malayalam, emotions are felt in the "liver," not in the "heart"). In fact, some of the cognates mean both "liver" and

"heart." Turner ascribes this to their common function of pumping blood through the body.

15. References

Bakker, Peter.

    1997   Romani and Dardic: Data from Fussman's Atlas. Report for the Romani Lexical Project, Aarhus University.

Dave, Jagdish.

    1995   Colloquial Gujarati: The Complete Course for Beginners. London/New York: Routledge.

Edelman, Dzhoi Iosifovna.

    1983   The Dardic and Nuristani Languages. Moscow: Nauka.

Grierson, George Abraham.

    1968a Linguistic Survey of India. Vol. 1, part 2, Comparative Vocabulary. Delhi: Motilal Banarsidass.

    1968b Linguistic Survey of India. Vol. 6, Indo-Aryan Languages, Mediate Group. Delhi: Motilal Banarsidass.

    1968c Linguistic Survey of India. Vol. 9, part 4, Pahari Languages and Gujuri. Delhi: Motilal Banarsidass.

    1968d Linguistic Survey of India. Vol. 11, "Gipsy" Languages. Delhi: Motilal Banarsidass.

    1969   The Piśāca Languages of North-Western India. 2nd ed. Delhi: Munshiram Manoharlal.

Hancock, Ian F.

    1995   A Handbook of Vlax Romani. Colombus, Ohio, US: Slavica Publishers,

              Inc.

    2000   The Emergence of Romani as a Koïné Outside of India. Scholarship

              and the Gypsy Struggle: Commitment to Romani Studies. Ed. by Thomas

              Acton. Hatfield, UK: University of Hertfordshire Press.

    2003   We Are the Romani People (*Ame Sam E Rromane Džene*). 2nd ed.

              Hatfield, UK: University of Hertfordshire Press.

    2006   On Romani Origins and Identity. University of Texas at Austin: Romani

              Archives and Documentation Center.

Jain, Sadhna Kantikumar.

    1984   *ChhattīsgaRhī aur khaRī bolī ke vyaakaraNõ kā thulnāthmak adhyayan* (A

              Comparative Study of the Grammars of Chhattisgarhi and Standard

              Hindi). 1st ed. Allahabad, India: Sahityavani.

Masica, Colin.

    1991   The Indo-Aryan Languages. Cambridge: Cambridge University Press.

Marsh, Adrian.

    2008   "No Promised Land"; History, Historiography & the Origins of the

              Gypsies. London: Greenwich University.

Sztojka Ferencz.

    1886   *Ö császari és magyar király Fensége József Föherczeg magyar és czigány*

              nyelv gyök-szótára. Románé álává. Kalocsa.

Tálos Endre.

    1999   Etymologica Zingarica. Acta Linguistica Hungarica. Vol. 46 (3-

            4), 251. Budapest: Akadémiai Kiadó.

Turner, Ralph Lilley.

    1924   *Traš*-, 'To Frighten.' Journey of the Gypsy Lore Society. 3rd series. Vol.

            III No. 1. Liverpool: Gypsy Lore Society.

    1969   A Comparative Dictionary of the Indo-Aryan Languages. London: Oxford

            University Press.

    2006   Later edition of the Dictionary, available at

            <http://dsal.uchicago.edu/dictionaries/soas/>.