
Syed Alamdar Nabi^{*}, Nasir Abbas^{**}, Firdos Atta^{***}
Munir Khan^{****}

On the Origin of Gypsies

ABSTRACT

Romani language, also called Gypsy, is a satellite language of the New Indo Aryan (NIA) family spoken mainly in Eastern Europe. The speakers of this language departed from the Subcontinent approximately one thousand years ago. They moved to Turkey via Iran and reached Eastern Europe. A very strong dominant view among the linguists is that Romani originated from Punjab and at the time of their departure from this area their forefathers either spoke some form of Punjabi or Hindi, whereas some researchers think that forefathers of the Gypsies were speakers of either Kashmiri or Sindhi or (Saraiki. In the current study, we highlight similarities between Gypsy (Romani) and Pakistani languages of NIA family and conclude that forefathers of the modern-day Romani were speakers of an older version of either Saraiki or Sindhi or Kashmiri. The core argument of this paper is based on the evidence that Kashmiri, Saraiki and Sindhi have palatalization and complex three-way pronominal suffixes with verbs which they share with Romani which is an indication of genetic relations between Romani and these Pakistani languages.

Keywords: Indo-Aryan, Romani, Saraiki, Sindhi, Grammar

* Assistant Professor, Government Emerson College, Multan, Pakistan

** Associate Professor, Department of English, Lasbela University of Agriculture Water and Marin Sciences, (LUAWMS), Uthal, Pakistan

*** Assistant Professor, Lasbela University of Agriculture Water and Marin Sciences, (LUAWMS), Uthal, Pakistan

**** Assistant Professor, Lasbela University of Agriculture Water and Marin Sciences, (LUAWMS), Uthal, Pakistan

Introduction

Romani is a language of Gypsy tribes called 'Roma' who are mainly clustered in Eastern European countries, although they are also sprinkled in other countries of the world. Their population in Europe was estimated 3.8 million in 2002 (Pan & Pfeil, 2003, p. 27). Roughly stated, there are 14 million Romani gypsies in the world. In the late nineteenth century, the scholars used the term 'gypsy languages' for scientific purpose. Later on the English term Romani was frequently used in literature. The origin of the Romani language has been a point of debate among the researchers working on Romani. Romani is classified by researchers as a satellite language of the Indo-Aryan family. Linguistically, Romani language had experienced plentiful external influences, over the centuries. The existing form of Romani is a blend of different Indo-European, Greek, Hungarian and Armenian languages. The presence of Indo-Aryan elements in Romani strengthens the view point of historians who had their place to the Subcontinent. The root cause of their migration is still a mystery but one could find an interesting debate in the literature regarding their exodus.

In the view of Hancock (2006) the forefathers of Romani speakers were recruited in the Army of an Afghan ruler, and therefore, they moved from their homeland for the sake of military adventures. Though previously, it was claimed that Romani speaking gypsies were musicians and were called 'Domes' in the Subcontinent (Matras, 1999a; Matras, 1999c). They moved from their homeland *as a result of* constant invasions of the Afghan Muslim invaders more than a thousand years ago.¹ The language of the Domes was initially called 'Domani' which the gypsies themselves later on changed into 'Romani' on reaching Greece when they realized the importance of Roman language and culture in the Eastern Europe (Matras, 1999). Matras considers that there is a very close genetic relation of Romani with the language of musicians called 'Domari' spoken in the Middle-East.

The current study is designed to elaborate the discussion on the origin of Romani. We highlight some linguistic connections between

¹The current paper focuses on only linguistic origin of the Romani speakers; therefore, we shall not analyze the arguments for and against the views on the main reason of gypsy migration. Interested readers are referred to John (2013) for a detailed discussion and analysis of both views.

Romani and North Western Indo-Aryan languages, particularly, Saraiki and Sindhi. The next section briefly elaborates previous research on the origin and background to the Romani language. The next part presents a comparative data of Romania and north western Indo-Aryan language, after analysis and discussion, the article is concluded in last section.

Background to the Gypsies

In this section we briefly recapitulate the views of researchers who have previously studied Romani with a view to determine its origin and classification. As the following review clearly indicates, almost all previous researchers agree that Romani originated in the Sub-continent and it is a language of Indo-Aryan family. However, the point of argument is to which particular branch of Indo-Aryan is Romani affiliated.

Earlier studies have diversified views on the origin of Gypsies. According to Hancock (2006), the forefathers of Romani gypsies migrated from the Subcontinent approximately one thousand years ago for military purposes. A major controversy between Hancock (2006) and Matras (1999b) is that the former asserts that the ancestors of the European gypsies were martial people and they moved from their origin for participation in wars but the latter claims that they were musicians and moved to Turkey and Greece because of wars. Almost six to seven hundred roots of Indo-Aryan words have been identified in Romani (Matras, 2011).

Although there is a consensus among linguists on the view that Romani is unquestionably an Indo-Aryan language, parallel claims about various sub-groups of Indo-Aryan family as a possible linguistic origin of Romani, have been forwarded. The disagreement is on the exact origin of Romani speakers. A dominant view is in favor of the Central Indo-Aryan (CIA) zone as the origin of Romani speakers. Urdu, Hindi, Punjabi, etc. are considered major representative languages of Central Indo-Aryan (CIA) group and Saraiki-Sindhi as those of North Western Indo-Aryan (NWIA) group. John (2013) identifies Romani as a Central IA language declaring it a closer relative of (Urdu)/Hindi and Punjabi. Genetic studies (Ali et al., 2009) also trace the origin of Romani population with Punjabi speakers. Turner (1926), Matras (2009) and Hancock (2006) argue that the Roma's ancestors spoke various Central Indo-Aryan languages. In other words, they also find the origin of Romani in Central Indo-Aryan.

In the opinion of Benišek (2020) Romani is classified in the Indo-Aryan family and suggested Romani place in Dardic branch of Indo-Aryan. His view is based on phonological similarities between Romani and Dardic languages like two sibilants, stop+rhotic and fricative+stop cluster. In the opinion of Benišek (2020, p.21), "Miklosich's view gained acceptance of the scholars of that time, e.g. Pischel (1883), Bloch (1919) who revised earlier opinion about the Eastern (Magadhi) origin of Romani (Grierson, 1888)." Thus Miklosich (1878) and Pischel (1883) were first to claim that Romani is a Dardic language by its origin. Zoller (2010) also attempts to establish relationship between Romani and Dardic languages of Northern areas of Pakistan on the basis of similarity between the two in lexicon and head-modifier compounds but his view faces criticism at the hands of Benišek (2020). He also points out that some of the analyses of Zoller are based on incorrect information

Turner (1926) is another serious attempt to identify the place of Romani in Indo-Aryan family. In this regard, Turner focused on changes which occurred in Indo-Aryan during transition from Old Indo-Aryan to Middle Indo-Aryan period and concluded that Romani belongs to Central Indo-Aryan sub-family. The analysis of Turner was also based on phonology. He considered development of rhotic vowel into full vowel, change of /kʂ/ cluster into aspirated velar /kʰ/ e.g. Sanskrit akṣi > Hindi ankʰ (eye), development of /mh/ into /m/ e.g. OIA asm > ahm > ham (Hindi), assimilation of stop+m into non-nasal labial e.g. atman > ap (Indo-Aryan *self*, /j/ affrication e.g. ja > dza (Hindi *go*) and spirantization of labial stops into /v/ e.g. grama > gav (village). These studies conclude that ancient Romani or what they call proto-Romani originated from central Indo-Aryan group of languages. However, Benišek (2020, p. 22) severely criticizes this view of Turner. Turner claims that forefathers of modern Romani speakers moved from India during the reign of the Mauryan emperor Asoka. According to Benišek, we find elements of post-Asoka Indian features in the speech of modern Romani (like change of /m/ > /v/) which confirm that forefathers of Romani were living in the Sub-continent even after Asokan reign. Benišek also points out that the common features that Turner identifies between Central Indo-Aryan and ancient Romani also exist in other languages of NWIA and even in the Eastern Indo-Aryan languages like Nepali (2020, p.24). Thus, this view is not so acceptable for researchers that Romani belongs to Central Indo-Aryan family or Hindi.

However, three things are important to keep in mind before we develop a final view about origin of Romani. These features which Turner finds common between Romani and central Indo-Aryan languages particularly Hindi, are also common between Romani and languages of North-western Indo-Aryan (NWIA). Hindi shares all words of Sanskrit listed above with Sindhi and Saraiki e.g. ankh, ham, gav, dza, etc. Along with this, Romani also shares vowel-ending in words, an additional linguistic characteristic which Romani shares with NWIA languages, e.g. Romani 'baero' (ship) berry (small boat) and 'poro' (read) are 'ḃero', 'Beṛi' and 'Porḃo' respectively last two of which are pronounced in Saraiki exactly as in Sindhi. Moreover, a common characteristic which Romani shares with NWIA (Sindhi and Saraiki) is retention of /r/ which central Indo-Aryan languages like Urdu, Punjabi and Hindi have lost. For example, from OIA tirin (three), 'draksha' (grapes), bhratr (brother), Hindi developed 'tin', 'dakh' and 'bhai' respectively but Saraiki has 'tre', 'drakh' and 'bhira' respectively from these cognates.

Thus, a very brief review of the previous studies on the origin of Romani clearly indicates that although a majority of linguists claim that Romani originated from Central Indo-Aryan and, therefore, the ancestors of modern Romani were speakers of a language which must be a sister of Punjabi or Urdu (or Hindi), but this claim is not thoroughly unchallenged. Latest studies have raised serious questions on the findings of the previous researchers like Matras and Hancock. A parallel view, which could not get much currency among the researchers, was of Beames and Sampson. Beames (1872) was also of the opinion that Romani is more closely related to Kashmiri than other Indo-Aryan languages; Sampson (1923) linked it with Multani (Saraiki) and Sindhi. These researchers, in a way, locate Romani in northwestern group of Indo-Aryan languages. In this study we extended this view by adding some similar linguistic data of Romani, Saraiki, Sindhi and some other Indo-Aryan languages.

Objectives of the Study

The main objective of the study is to find out the linguistic connections between Romani and Indo-Aryan languages. Moreover, the study also highlights the differences and similarities between Romani, and Pakistani languages especially Sindhi, Saraiki and Kashmiri to determine its relationship with these languages. In this way, we may determine the origin of the Romani language.

Methodology

To investigate the origin of the Romani, a comparative analytic approach is adopted in this study. This study highlights the existing similarities between North-Western (Sindhi, Saraiki, Kashmiri, etc.) languages of Indo-Aryan (IA) sub-family and various dialects of Romani to determine whether the forefathers of Romani migrated from the area where Sindhi and/or Saraiki languages were spoken at that time or alternatively, they migrated from the areas where Punjabi and Hindi were being spoken. Many linguistic features which may evolve or eliminate diachronically, are examined to compare the pattern of Romani with IA. Further, a comparison of Romani with Kashmiri is also part of the discussion section. On the basis of these processes of data analyses, the conclusion about the origin of Romani is drawn.

Status of Romani in Indo-Aryan

In the last quarter of the 18th Century CE, the above European linguists had realized that the language of Gypsies of the East European countries was very closely related to Indo-Aryan branch of Indo-Iranian family which itself was a branch of Indo-European family of languages. A lot of similarities between Romani and Indo-Aryan languages have been found in the previous research, which is briefly summarized in the previous section, which highlights close genetic relation of Romani with languages of the Sub-continent. In the following points we briefly enumerate some major linguistic characteristics which Romani shares with the Indo-Aryan (IA) languages;

- i) Agreement of verb and adjectives in terms of number and gender is a very common phenomenon of Indo-Aryan languages. Similarly, number and gender agreement of nouns with verbs and adjectives does exist in Romani (Matras, 1999b).
- ii) A very prominent Indo-Aryan feature of Romani is ergativity. Ergativity is a very characteristic feature of IA languages and the same feature is also found in Romani language.
- iii) Another important feature of Romani which confirms its genetic relation with IA languages is pro-drop. Pro-drop is very common in IA languages. Romani speech exhibits pro-drop (Matras, 2011).

- iv) At phonological level, similarities between Romani and IA languages do exist. One such feature is existence of aspirated palatal affricates in Romani and IA languages.
- v) Contrastive Vowel length is a shared prominent feature of some dialects of Romani and IA languages.
- vi) Romani is an SOV language with ample flexibility allowed in word order. Indo-Aryan languages are also SOV languages and allow change in word-order.
- vii) Another important characteristic of Romani which shows its similarity with IA family is the prefix *bi-* which is used to derive antonyms and used in the meaning of 'without' (Matras, 2004, p. 74).
- viii) Adding a vowel suffix 'a' with words ending on vowels to form a vocative is common in languages of IA family. The same is also found in Romani as the following expressions illustrate.

Romani: p^hurea! (O old lady!)

Hindi/Urdu: b^huṛia! (O old lady!)

- ix) Some dialects of Romani still use the common expression of the language of Lucknow 'amã' as a conjunction pronounced without nasalization at the end of the word i.e./ama/ (Matras, 2004, p.94). Romani, like Hindi/Urdu, also uses pronouns 'amaro' and 'tumaro' for possessive forms of first person plural and second person respectively (Matras, 2004, p.76). This is a duplication of 'hamara' and 'tumhara' of Urdu/Hindi respectively, which came into being as a result of [h] deletion, a phenomenon common in the world languages (Mielke, 2002).
- x) Substitution of [v] to [b] is a common feature of IA languages. In the historical development of IA, a kind of fortition or hardening occurred and some languages of the Sub-continent like Urdu/Hindi substituted [b] with [v] in the onset of words of Sanskrit Vedic origin (Masica, 1991). But some other languages of the same family retained the original labio-dental approximant. The following examples in (1) illustrate both cases.²

² Si=Sindhi, Sr=Saraiki, U/H=Urdu/Hindi

(1)

[b]-languages	Glosses	[u]-languages
Bas	(capacity)	uas (Si, Sr, P)
Bars	(year)	uarsh (U/H)
Bars	(year)	uareh (P)
Bi:s	(twenty)	ui:h (Sr)
Betho	(sit)	ueh (Si)
Baras	(rain)	uas (Sr)

The above data show that IA languages may be categorized into b-languages and u-languages based on this particularly phenomenon. On the same pattern, Matras (2004, p.70) describes a free variation between [u] and [b] in different dialects of Romani.

- xi) In the study of historical development of IA, we come across examples of substitution of [h] with [s] (Masica, 1991). This phenomenon is also common in other world languages (Mielke, 2002). In the following data in (2), we provide examples of such substitution from IA family of languages;

(2)

[s]-languages	Glosses	[h]-languages
sans (U)	(breathing)	sah (P, Sr)
phans (U)	(gallows)	phah (Sr)
ghas (U)	(grass)	gha (Sr) (Persian 'kah')
kapas (U)	(cotton)	kapah (P/Sr)

In the words of Matras (2004), consonants [s] and [h] are also interchangeable in various dialects of Romani.

- xii) Free variation between [s] and [ʃ] is very common in various languages of IA family particularly those spoken in India. This is actually an influence of three alveolar fricatives in Vedic Sanskrit which later on merged into two fricatives in modern day IA languages. Some speakers adopted one phoneme and others adopted the other. Pashto still has retained retroflex fricative along with non-retroflex anterior and posterior sibilants (David, 2014, p. 9) which reminds us of the influence of Sanskrit or Proto-Indo-Aryan on Pashto. Owing to such historical reasons, substitution between [s] and [ʃ] is frequent in the speech of speakers of IA languages in India. Matras (2004) is of the opinion that such free variation between [s] and [ʃ] is also found in Romani.

- xiii) Insertion of a lateral consonant [l] in words is also seen in languages of Indo-Aryan family. Examples of this phenomenon are seen in the Urdu/Hindi words, etc. Similarly, we find a frequent l-insertion in words of Romani. Some examples of this phenomenon in Urdu/Hindi are listed in the data set (3) in the following section.

(3)

Words	L-insertion	Glosses
tak	talak	(till)
dikha	dikhla	(expose)
'dikha	dikhal(Sr)	(expose),
bhora>	bhoral (Sr)	(whitish)

- xiv) Deletion of consonant [h] is common in IA languages. Saraiki speakers do the same with words of Indian origin. e.g. OIA *kōmh*, *khūha*, >*kumāvaṇ*, *khūhin* Saraikibut *kumhlānā*, *khūh* (wither, well) in old Hindi respectively. Similar h-deletion also occurred in Romani in historical process parallel to that in IA languages. Romani pronouns 'amaro' and 'tumarō' were actually derived as a result of h-deletion on word-initial position. Such Romani words as 'ova' (*That one*, <Saraiki 'oha'), 'but' (*very*< IA 'bahut') also remind us of the diachronic process of h-deletion.
- xv) In connection with the above, it is important to note that in diachronic development, Romani seems to have passed from a stage when open margins were not allowed. Masica (1991) is of the opinion that IA languages also passed from such a stage in the past. The repair strategy to satisfy such a constraint was consonant insertion. Romani seems to insert approximant [ʋ] in such cases, which Matras (2004, p. 99) calls ʋ-prothesis in Romani dialects. A long list of words of Romani starting with such an approximant (which the other Indo-Aryan cognates lack) indicates this historical development. For example, in Romani, words for 'lips' and 'hand' are 'vost' and 'vast,' respectively. Both are derived from Indo-Aryan 'host' and 'hast' respectively and the path of their development is that first h-deletion occurred giving 'oth' and 'ast' which later on became 'vost' and 'vast' as a result of ʋ-prothesis. In some IA languages, insertion of [ʋ] is very common in morphological constructions. We derive causatives in Urdu, Hindi and Saraiki, etc. by inserting [ʋ] to break hiatus of vowels. The examples are 'kar>kara>karva, etc. Saraiki uses this epenthesis for getting double causatives which are discussed in some detail in the following section.

xvi) Beside, these structural similarities, Romani shores a lot of words with modern IA languages. Some of those words are reproduced in (4) below from (Mustafa, 2019);

(4)

Words	Glosses
mang	beg
bal	hair
kan	ear
gi	soul, heart
jan	know
pani	water
dukh	pain
gin	count
rani	lady
dori	thread
loon	sault
dar	fear
rat	night

xvii) Loss of retroflex sounds occurred in some of the Indo-Aryan languages in historical development from Middle Indo-Aryan to New Indo-Aryan (Masica, 1991). Some languages of the family substituted retroflex sounds with the corresponding non-retroflex sounds but others retained the original retroflex consonants and some others substituted these consonants partially. For example, Hindi/Urdu lacks nasal retroflex which Punjabi, Sindhi and Saraiki have retained. However, further distinction between Saraiki and Punjabi is that the former uses retroflex and non-retroflex nasals for phonemic contrast but the later, in most of its common dialects, uses these two types of consonants as dialectal variants positing them at allophonic level in Punjabi. Romani has thoroughly lost the retroflex consonants. The loss of retroflex sounds in Romani is illustrated in the following data in (5);

(5)

Pakistan³	Romani	glosses
Pāṇī	pani	water
ʃaṇ	jan	know
gīṇ	gin	count
lūṇ	loon	salt
suṇ	sun	listen
baṛi	bari	big
beṛi	berri	small ship

xviii) During historical development, breathy voiced sounds of IA languages substituted with either plain voiced or voiceless aspirated sounds in Punjabi (Masica, 1991) and Kashmiri (Koul & Wali, 2006). The following examples given in (6) show this;

(6)

Pakistan	Romani	glosses
ghar	kaeri < kher	home
dho	Tov < Thov/thov	wash
dhú:	Tuv < Thuv	smoke

In some of the cases, Romani substituted words containing breathy voiced segments with those containing plain voiced or voiceless but in others, it substituted breathy voiced segments with voiceless aspirated ones like Punjabi and Kashmiri.

xix) Palatalization is common Saraiki (Atta, 2019). For deriving past tense and plurals, the stem-final consonant is palatalized. Some examples of Palatalization in Saraiki are given in (7) below;

(7)

phasi (entangle TR V SG F) phas+ia > phsʃa (Past V PL F)
rukki (stop INT V SG F) rukk+ĩa > rukkiĩ (Past V PL F)
varri (ripe INT V SG F) varr+iĩ > varriĩ (Past V PL F)
kali (Black F SG N) kali+ã > kalĩã (PL N F)
nali (Channel F SG N) nali+ẽ > nalĩẽ (Obl. PL. N F)

Palatalization is also found in Kashmiri (Bhaskararao et al., 2009). Romani also exhibits palatalization in form of insertion of a palatal segment which occurs in past tense (Matras, 2011). In the following paragraphs several examples of palatalization in Romani are quoted. The words like Išom/ išjom (I am), panjsa (with water), phenja (sister),

³ These words are spoken in Saraiki, Sindhi and Punjabi along with Kashmiri in Pakistani languages.

etc. are examples of palatalization in Romani quoted in Matras (2004).

- xx) Another important characteristic change in historical development of IA in most of the languages of the family is metathesis of s/h which results in secondary articulation. This displacement of a segment and its transformation to a secondary articulation on the landing site occurred in the period when Old Indo-Aryan (OIA: Vedic Sanskrit) was changing into Middle Indo-Aryan (MIA). First [s] changed into [h] and later on it moved to the landing place as a secondary articulation. Thus, the development was as suggested by (Kobayashi, 2004), st>ht>th>t^h

The following examples of Sanskrit taken from Turner (1966) show this process.

(8)

Sanskrit	Saraiki	Romani⁴	Glosses
vasati	bet ^h a	besto	sit
haṣṭa	fiṭ ^h	vast	hand
stambha	thamb ^h a	thãbla	pillar
stabdha (firmly fixed)	t ^h add ^h a (cold) Saraiki	tardo	upright
stana	t ^h añ	—	udder
stanya	t ^h añ	—	milk
stoka	t ^h oṛa	—	little

In these data, Romani examples show that it has retained the words with 'st' clusters particularly on word-final position. It simply means that the development which occurred in NIA languages did not occur in Romani and the lexicon of Roman still retains the old versions of these words.

- xxi) In Sindhi, an epenthetic vowel is added as a singular marker at the end of close syllables, a practice which is quite common and frequent in Sindhi, though subtly diminishing in urban areas of Sindh (Buglio, 1993). Similarly, in some varieties of Romani, some words still retain such a vocalic ending which reminds us of the past history and origin of the Romani gypsies. Some examples in (9) from Romani compared with those in Sindhi illustrate this fact;

⁴ The last three words could not be found in Romani.

(9)

Romani	Glosses	Sindhi
Baro	(big)	vado
vando	(egg)	anu:
učo	(high)	učo
kando	(thorn)	kando

Similarly, we also note a word-final short vowel in the words ending on stops in Saraiki. This is a common phenomenon in rural Saraiki.

All these examples quoted above reveal enough evidence to substantiate that Romani is a language of IA family. In the following section, we analyze these data to identify the origin of Romani in the IA family of languages.

Analysis and Discussion

The above similarities between Romani and IA languages are quite clear to support the claim of Matras (1991) and others that Romani is an Indo-Aryan language. Here the point of discussion is the status of Romani within Indo-Aryan. With reference to the above similar characteristics of Romani and IA languages, below the table clarifies to which IA language the Romani resembles most.

Table 1: Language-wise List of similar features which IA languages share with Romani ⁵

S. No.	Features in Romani	Sindhi	Saraiki	Kashmiri	Punjabi	Urdu
1	No Breathy voicing	X	X	✓	✓	X
2	Agreement (number,gender)	✓	✓	✓	✓	✓
3	Vowel ending	✓	✓	X	X	X
4	Ergativity	✓	✓	✓	✓	✓
5	h-metathesis	✓	✓	✓	✓	✓
6	Aspirated affricates	✓	✓	✓	✓	✓
7	Vowel length contrast	✓	✓	✓	✓	✓
8	SOV structure	✓	✓	✓	✓	✓

⁵ In Table 1, some of the features like pronominal suffixation and palatalization listed in the table which have not been discussed in the previous section are discussed in detail in section 4 below.

S. No.	Features in Romani	Sindhi	Saraiki	Kashmiri	Punjabi	Urdu
9	Prefix bi-	✓	✓	✓	✓	✓
10	Loss of retroflex	X	X	X	X	X
11	Vowel ending	✓	✓	X	X	X
12	Palatalization	X	✓	✓	X	X
13	Double causative	X	✓	X	X	X
14	3-way pronominal suffixes	✓	✓	✓	X	X
15	2-way pronominal suffixes	✓	✓	✓	X	X
16	1-way pronominal suffixes	✓	✓	✓	✓	X
	Total	12	14	12	9	7

In view of above table, Saraiki, Sindhi and Kashmiri seem to have maximum number of common linguistic features with Romani as compared to Punjabi and Urdu (and Hindi) which share lesser number of features with it. A major structural similarity between Punjabi and Romani is that both have lost breathy voiced segments from their phonemic inventory. However, this may not be the *only* decisive factor in the current context because of two reasons. First, Kashmiri also shares the same feature with Romani as it also does not have breathy voiced segments. Thus, in this regard, Kashmiri equally qualifies for claiming genetic relation with Romani. Secondly, breathy voiced segments are a very difficult phenomenon in the world languages. The similarity between the two languages is on account of absence of a very marked feature. This similarity may also be an indicator of similar independent direction of development in two languages according to the principles of universal markedness. Many languages of the world do not have breathy voiced consonants. Since these consonants are very difficult, it is natural if any two languages delete these consonants independently of each other. Thus, presence (not absence) of a particular common feature in two languages may be considered a stronger indicator of genetic relations between two languages.

In the same line, we see similarity between Romani and NWIA languages (Saraiki-Sindhi-Kashmiri) in terms of a common feature which is three-way pronominal suffixation. Romani like these three NWIA languages (namely Kashmiri, Saraiki and Sindhi) has three-way pronominal suffixation, in that, it has pronouns of subject, object separately and together (subject and object) suffixed with verbs. Existence of pronominal suffixes in Kashmiri, Sindhi and Saraiki are very frequent. This phenomenon exists in other languages of this family like Khetrani (Birmani & Ahmed, 2017) and is most common in Persian language.

Because of this, some Oriental researchers have ascribed Saraiki and Sindhi pronominal suffixes to Persian influence (Emeneau, 1980). However, some other researchers had strongly disagreed to this view and argued that pronominal suffixes in Sindhi, Lahnda and Kashmiri are inherently an Indo-Aryan phenomenon (Grierson, 1895). The following examples in (10) illustrate this common phenomenon in Romani, Sindhi and Saraiki;

(10)

Romani	Glosses	Saraiki	Sindhi
Geljom	(I went)	gijum	vioṣ/vias(F)
dikhlom	(I saw)	di t ^h um	di t um
kerdjom	(I did)	ki: t ^h um	keum
aljom	(I came)	ajum	ajos/ajaṣ (F)
mokhlis	(he left)	mokl ^h is	moklajai
puchljom	(I asked)	puc ^h um	puc ^h um

Punjabi (Butt, 2007) also exhibits this characteristic in the form of one-way suffixation or cliticization, whereas Urdu/Hindi lacks this. However, the important aspect of this grammatical phenomenon in the Western Indo-Aryan languages (Saraiki, Sindhi, Khetrani, Jaffarki and Kashmiri) is that they attach pronominal suffixes of subject and object with verb which is exclusive feature of this sub-family. Some examples of double suffixation of pronouns from Romani are reproduced below in (11) from Matras (2011, 2004);

(11)

Romani	Glosses
Kard-os-is	He did it
kardomis	I did it
dikhljias	he saw me
'dikhljomos'	I saw him
'dikhljomi'	I saw her

Similarly, suffixation of pronouns of subject and object is also found in Kashmiri and Sindhi. Some such examples of such double suffixation are noted below in (12);

(12)

A.Sindhi

disandias
disandos
dit^hogn
dit^haun
dit^hoin
dit^hain
dit^hosin

Glosses

I (F.) will see him/her
I (M.) will see him/her
I saw them
you (plural) saw them
you (sg) saw them
s/he saw them
We saw them

B. Kashmiri

morimava
balravak
dopnas

Glosses

I killed you
I will heal them
he spoke to him

(Grierson 1895, p.341).

tsichu-h-an vucha:n

You be-2.Sg.NOM-3.Sg.ACC see

You are looking at it (Sharma, 2001, p. 226)

The following examples in (14) from Saraiki confirm the existence of this phenomenon in Saraiki.

(14)

akh> (akh+ie+m+is) >akhiemis

I asked him

saṭ> (saṭ+es+n+is) >saṭesnis

They will discard him

mar>(mar+ie+m+is) >mariemis

I have hit him

Because of spatial constraints, only small number of examples has been written here. Almost all transitive verbs of Saraiki can have such double suffixation. Three-way suffixation is a very complex phenomenon. It also exists in Sanskrit (Bubenik, 1992, p. 8) which is a very convincing indicator that this is an inherent Indo-Aryan phenomenon. Urdu/Hindi and Punjabi do not have such suffixation but Romani does have it. We cannot expect that the same phenomenon had disappeared from the language of the forefathers of Romani speakers at the time of their migration but the same phenomenon re-emerged in Romani on reaching Europe. The more probable interpretation is that Romani speakers migrated from an area where the language of the people had such complex suffixation. Such a language or family of languages can be that

comprising of Sindhi, Saraiki, Kashmiri, etc. Because of these features Grierson (1895) has placed Kashmiri in the same sub-family i.e. North-Western Indo-Aryan with Saraiki and Sindhi.⁶

Another important feature which strengthens our view is that Saraiki shares with Romani the double causatives. In Saraiki a morpheme 'a' is added with stems of verbs to get the first causative and again another morpheme 'a' immediately preceded by an epenthetic approximant 'v' is used to derive double causative form, e.g. 'kar' (do) 'kar+a>kara' 'help/hand to do' and 'kar+va>karva' 'to get it done by someone else'. Similar pattern is also seen in Romani in which causative verbs are derived from transitive verbs or nouns by adding semi-vowels e.g. Romani 'dar-av-av' (to cause someone to frighten) (Matras, 2011, p. 262) (vis-a-vis Saraiki double causative 'darvavan'). The same feature also exists in Urdu and Hindi but not at such a large scale as it does in Saraiki.

In the above discussion, we have discussed that Romani shares more features with Saraiki, Sindhi and Kashmiri than with Punjabi, Urdu and Hindi. We know that Romani speaking gypsies migrated from the sub-continent one thousand years ago. At that time Multan was a part of Sindh and the language of the area around Multan was considered one of the dialects of Sindhi. Since wars waged by Afghan invaders were the major causes of the migration of gypsies, it is quite plausible to assume that the forefathers of these gypsies were the inhabitant of those areas which were major battlegrounds and routes of Afghan armies of that time. It is almost the same area where old Indus valley languages also called North Western Indo-Aryan languages (Sindhi and Multani) and Dardic languages meet. The isogloss these two sub-families lies on the northern border of Multani language, keeping in view that the borders of Multan of that time extended to the farther north from where it is confined these days. It is also important to note that Punjabi has also geographically advanced westwards from where it was one thousand years ago when Gypsy migration took place. These facts support the view that the ancestors of Romani speakers migrated from an area where Multani (an older name of Saraiki) was spoken at the time of their departure.

Finally, we try to address the question that why Western linguists could not realize this. The reason for this is that the Western linguists first took Hindi and later on Punjabi the major representative of IA family of

⁶ Both expressions, North-Western Indo-Aryan and Western Indo-Aryan, have been used interchangeably in the literature.

languages. Since, Romani is also an IA language, they found several similarities between Punjabi and Romani (some of those are listed above) which convinced them of the accuracy of their point of view. They also noticed the existence of palatalization and double pronominal suffixation in Romani but as they did not see similar features in either Hindi (Urdu) or Punjabi, they ascribed these to some European languages. For example, Matras (2004) thinks that Romani took palatalization from Greek and double suffixation either from Italian or from Domari (Matras, 1999) (without further commenting on where the Domari took this feature from). However, it is more convincing to assume that these are genetic features of Romani which, along with Saraiki and Sindhi, it inherits from its ancestors. Languages lose difficult or marked features and take easier or unmarked features as loan as a result of social contact with other neighboring languages. Thus, loss of breathy voiced and retroflex consonants can be an artifact of deletion as a result of diachronic development in Romani as well as Punjabi and Urdu but such difficult/marked features as double causative and double pronominal suffixation are genetic features of Romani which it shares with its sister languages spoken in Pakistan.

Conclusion

In this paper, we highlighted similarities between Romani and IA languages to establish that Romani is a language of Indo-Aryan family. Later on, we highlighted similarities among NWIA and Central Indo-Aryan languages with Romani and reached a conclusion that Romani is genetically and structurally closer to Sindhi, Saraiki and Kashmiri. On the basis of this, we conclude that forefathers of Romani speakers migrated from an area where an older version of a dialect of the NWIA sub-family was spoken. The language of the people of that area was called Sindhi or Multani at that time. Our conclusion is supported by the fact that a group of Romani are also called Sinti (Setti, 2017), an indicator that Romani speakers have retained their old name 'Sindhi' with the loss of breathy voiced segment. The genetic similarity found between modern day Romani speakers and people of Punjab (Ali et al., 2009) further confirms this view. The linguistic variation in those areas is a result of language replacement and development. However, this is a working hypothesis that may be further investigated, challenged or supported with empirical evidence.

References

- Ali, M., Martin, M., Adam, B., Parry, D. A., & Jain, P. (2009). Null Mutations in LTBP2 Cause Primary Congenital Glaucoma. *American Journal of Human Genetics*, 84(5), 664-671.
- Atta, F. (2019). *Phonetics and Phonology of the Saraiki language: a descriptive exploration and an analysis from the perspective of Optimality Theory* (PhD dissertation), Shanghai: International Studies University, Shanghai.
- Beames, J. (1872). *A Comparative Grammar of the Modern Aryan Languages of India*. London: Trübner & Co.
- Benišek, M. (2020). The Historical Origin of Romani. In Y. Matros & A. Tenser (eds.) *The Palgrave Handbook of Romani Language and Linguistics* (pp.13-48). Manchester: Macmillan.
- Bhaskararao, P., Hassan, S., Naikoo, I. A., Ganai, P. A., Wani, N. H., & Ahmad, T. (2009). A phonetic study of Kashmiri palatalization. *Working Papers in Corpus-based Linguistics and Language Education*, 3, 1-17.
- Birmani, A. H., & Ahmed, F. (2017). Language of the Khetrans of Barkhan of Pakistani Balochistan: A preliminary description. *Lingua*, 191, 3-21.
- Bloch, J. (1919). *La formation de la langue marathe*. Paris: Edouard Champion.
- Bubenik, V. (1992). On the use of pronominal clitics in late middle Indo-Aryan. *Vienna Journal of South Asian Studies*, 36, 7-18.
- Bughio, M. Q. (1993). *A comparative sociolinguistic study of urban and rural Sindhi : (a study of phonological variation in Vicholi Sindhi, Pakistan*. (PhD), University of Essex, UK.

- Butt, M. (2007). The role of pronominal suffixes in Punjabi. In Z. e. al. (Ed.), *Architectures, rules and preferences: Variations on themes* (pp 341-368). Bloomington, Indiana: CSLI Publications.
- David, A. B. (2014). *Descriptive grammar of Pashto and its dialects* (Vol. 1). Berlin: Walter de Gruyter.
- Emeneau, M. (1980). India and Linguistic areas. In A. S. Dil (Ed.), *Languages and Linguistic Area: Essays by Murray B. Emeneau* (pp. 126-166). California Stanford University Press.
- Grierson, G. (1888). Doms, Jats, and the Origin of the Gypsies. *Romani Studies*, 1(2), 71-76.
- Grierson, G. (1895). On pronominal suffixes in the Kashmiri language. *Journal of the Asiatic Society of Bengal*, 64(4), 336-351.
- Hancock, I. (2006). On Romany origins and identity: Questions for discussion. In A. Marsh & E. Strand (Eds.), *Gypsies and the problem of identities: Contextual, constructed and contested*(pp 69-92). Istanbul: Swedish Research Institute.
- John, V. G. (2013). *A re-examination of the origins of Romani*. (PhD), University of Texas.
- Kobayashi, M. (2004). *Historical Phonology of Old Indo-Aryan Consonants*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa Tokyo University.
- Koul, O. N., & Wali, K. (2006). *Modern Kashmiri Grammar*. Springfield: Dunwoody Press
- Masica, C. P. (1991). *The Indo-Aryan Languages*: Cambridge University Press.
- Matras, Y. (1999a). The state of Domari in present-day Jerusalem. *Mediterranean Language Review*, 11, 1-58.

- Matras, Y. (1999b). Subject Clitics in Sinti. *Acta Linguistica Hungarica*, 46(3-4), 147-168.
- Matras, Y. (1999c). Writing Romani: The pragmatics of codification in a stateless language. *Applied Linguistics*, 20(4), 481-502.
- Matras, Y. (2004). Romacilikanes -- The Romani Dialect of Parakalamos. *Romani Studies*, 14(1), 50-109.
- Matras, Y. (2009). Romani. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (pp. 898-900). Oxford: Elsevier.
- Matras, Y. (2011). Romani. In B. Kortmann & J. v. d. Auwera (Eds.), *The languages and linguistics of Europe. A comprehensive guide* (pp. 257-268). Berlin: De Gruyter Mouton.
- Mielke, J. (2002). Turkish h-deletion: Evidence for Interplay of Speech Perception and Phonology. *ZAS papers in linguistics*, 28, 55-72.
- Miklosich, F. (1878). *Beiträge zur Kenntnis der Zigeunermundarten* (Vol. 4). Vienna: Gerold.
- Mustafa, G. (2019). *A comparative study of phonological development in Lasi and Romani*. (MS Dissertation), LUAWMS, Uthal Balochistan.
- Pan, C., & Pfeil, B. S. (2003). *National Minorities in Europe: Handbook*. Vienna: Braumüller.
- Pischel, R. (1883). Die Heimath der Zigeuner. *Deutsche Rundschau*, 36, 353-375.
- Sampson, J. (1923). On the origin and early migration of the Gypsies. *Journal of the Gypsy Lore Society*, II(4), 156-169.
- Setti, F. (2017). Long-lasting fieldwork, ethnographic restitution and 'engaged anthropology' in Romani studies. *The Urban Review*, 49(3), 372-381.

Turner, R. L. (2007). Urdu. In George Cardona & Danesh Jain (editors), *The Indo-Aryan Languages* (pp.286-350). Londo: Routledge.

Turner, R. L. (1926). The position of Romani in indo-Aryan. *Journal of the Gypsy Lore Society* 5, 145-189.

Turner, R. L. (1966). *A comparative dictionary of the Indo-Aryan languages*. London: SOAS.

Zoller, C. P. (2010). Aspects of the early history of Romani. *Acta Orientalia*, 71, 70-70.