The Origins of Romani

- Valyi István 1760 Hungarian Calvinist theology student in Leiden he discussed Sanskrit with three visiting Malabar students. He recognized similarities between Sanskrit (which they used liturgically) and the little bit of Romani that he knew.
- This story was published and promoted by Jacob Rüdiger in 1782. In 1785 William Marden published a comparative list of words in English Romani, Turkish Romani, and Hindi. He concluded:

"... Should any doubt ... of the identity of the *Gypsey* or *Cingari*, and the *Hindustanic* languages, ... tribes wandering through the mountains of Nubia, or the plains of Romania, have been conversing for centuries in a dialect precisely similar to that spoken at this day by the obscure, despised and wretched people in England, whose language has been considered as a fabricated gibberish and confounded with a cant in use among thieves and beggars, and whose persons have been ... an object of the persecution, instead of protection, of our laws."

- August Pott (1844) the first serious comparative study of Romani. He concluded that all dialects of Romani derive from a single language. He also noted that Romani was a language of its own, and not a version of thieve's jargon. Finally, he proposed that Romani is most closely related to Sindhi.
- Later proposals related Romani to Dardic (NW Indo-Iranian, Franz Mikosich, 1874 and 1878) or Eastern Indic (George Grierson, 1903-1922).
- John Sampson (1923) proposed that Gypsies entered Persia in the 9th century, speaking a North Western Indic language while Dardic had devoiced Indic voiced aspirates, not all dialects of Romani had:

(1) Sanskrit	European Romani	Lomavren	Dome	ari
ghāsa dhāv-	khas thov-	khas thov-	gas dau	'hay' 'wash'
bhrātr	phral	phal	bar	'brother'

• Ralph Turner (1926) disputed the NW origin, and proposed a central Indic affiliation. He followed Sampson in dividing Romani into three branches: European, Armenian, and Syrian. (Romani, Lomavren, and Domari) - Turner suspected the split occurred before leaving India.

Three Dialects of Romani

(2)	Sanskrit		Domari	i	Lomavren		Romani	
	t atta	:	t atos	:	r arav	:	r varo	'flour'
	bh bhusám	:	b bis	:	ph phus	:	ph phus	'straw'
	v vla	:	w wal	:	v valin	:	b bal	'hair'
	h bahu	:	h bahut	:	h bahu	:	Ø but	'much'
	m nāma	:	m nam	:	m nam	:		'name'
	š širas	:	s siri	:	S SĪS	:	š šero	'head'
	d doma	:	d dom	:	l lom	:	r rom	'man, husband'

- Romani and Domari differ in another important way they have different loanword vocabularies:
- Romani has several loanwords from Iranian (eg. *baxt* 'luck'), Armenian (*grast* 'horse'), and Greek (*drom* 'road').
- Domari has loans from Persian (e.g. *kangri* 'wagon') and Arabic (a large part of the vocabulary).
- Of their Persian loanwords, Romani and Domari share only about 10-15% (Hancock 1995).

The Origins of the Gitanos in Spain

From the North

- The earliest reference of Gitanos in Spain is from a 1415 letter of passage by (the future) Alfonso V for Tomás de Sabba, a pilgrim to Santiago de Compostela.
- The more commonly cited date is 1425, when Alfonso V issued a letter of passage in Zaragoza to Juan de Egipto Menor ('Little Egypt') (López de Meneses 1968, 1971).
- This self-reference to Egypt is the source of the ethnic label *Gitano* (< *Egiptano* 'Egyptian').
- Similar self-designation led to the terms *Gypsy* in England and *Gitan* in France.
- Roma who arrived in these countries in the 14th-16th centuries said they were pilgrims from 'Little Egypt' (Fraser 1992, ch. 4).
- The legend of Egyptian origin remains widespread and is not unique to Spain.

From the East

- A separate group of Gitanos arrived in Spain from Greece, via the Mediterranean coast, beginning around 1480; the first attested arrival was Juan de la Costa, who claimed to be fleeing the Ottomans (López de Meneses 1968).
- Contrast between 'Egyptians' and 'Greeks' were noted by Pedro Salazar de Mendoza, who, in 1618, commented on differences in dress and occupations between the two groups (Leblon 1994:15).
- These two attested routes, from the North and from the East, are consistent, both geographically and chronologically, with the general diaspora of Roma between the 11th and 15th centuries.

From the South?

- The NORTHERN AFRICA HYPOTHESIS: An alternative hypothesis, which, while commonly cited, lacks historical attestation. This view holds that Gitanos first entered Spain from the South, via Northern Africa.
- There have been a few arguments for this view, but the most common reason has to do with the claim of Egyptian origin:

"The oldest document stating this migrating wave, signed by Alfonso The Magnanimous, dates back to 1425. The possibility that they penetrated before, via Africa, is not ruled out, this would support the theory that they came from Egypt and would also explain the etymological origin of the word." <u>http://flamenco-world.com/magazine/about/historia_del_flamenco/paginas/8.htm</u>

• Other arguments are from unnamed authorities; some even claim historical records, although these are never named:

"There are records of their having arrived in Spain from Northern Africa, as early as 1425 ..." <u>http://en.wikipedia.org/wiki/Gitanos</u>

"There are authors who affirm that certain tribes already arrived in Andalusia from Yeman in 755, following the armies of Abderramán. ... The Andalusian Gypsies, then would be called 'Egyptians', and those from the North 'Greeks'." Manuel Herrera Rodas, Prologue to Zoido Naranjo (1999: 13&15) [my translation]

- Clébert (1963) is often cited in connection with the Northern Africa Hypothesis. He suggests that linguistic and cultural differences point to different histories between Gitanos and other European Roma.
- Clébert also cites Serboianu (1930), who suggests that Gitanos belong the group of Nubian and Egyptian Gypsies who were found in Crete in 1422 this accounts for a 1540 report that Gitanos could speak Greek.

Alternative Explanation:

- Stories of Egyptian origin are widespread, even among Roma elsewhere in Europe.
- Linguistic and cultural differences may point to different histories, but these differences could be due to the attempts of forced assimilation that Gitanos underwent in Spain between the 16th and 18th centuries (Leblon 1994).
- The most likely account for the fact that some spoke Greek in 1540 would be that they had come from Greek-speaking areas, as many had already noted.

Mis-interpreting Sampson

• San Román (1976) provides another basis for the Northern Africa Hypothesis, based on an interpretation of Sampson's 1926 classification of Gypsy languages:

"Analyzing the dialects of Romani, ..., he reconstructed the route the Gypsies followed from the East to Europe. Leaving India, they passed through Afghanistan to Iran. From Iran, ..., they took two different paths: some arrived in Byzantium by way of Armenia, while others passed to Northern Africa, having crossed Syria." (p. 18 – my translation)

- While Sampson did propose two branches of Romani 'European' and 'Asiatic', he did not mention the Asiatic branch west of Egypt, nor did he propose that Gitanos were related to this Asiatic branch.
- Sampson's two groups are now commonly referred to as *Roma* and *Dom*.
- Roma are represented by several European subgroups, including Balkan, Vlax (Romania), Central (Czech, Slovak, Hungary), Northern (Germany, France, Russia, Poland), British, and Iberian (Matras 2002).
- Dom are Middle Eastern, including the *Nawar* and *Kurbát* (Syria, Palestine), *Helebis* (Egypt), and *Karači* (Persia, Asia Minor).
- Although Sampson proposed these groups diverged sometime after an initial migration to Persia, others have argued the Roma and Dom represent separate migrations from India (e.g. Turner 1926, Hancock 1995)
- San Román's suggestion does suggest a consequence of the Northern Africa Hypothesis. Given that the Roma took a more Northern route into Europe and the Dom occupied the Middle East, if some Gitanos came to Spain via Northern Africa, they would be Dom, not Roma.

- Because we have clear attestations of Roma entering Spain for the North and East, the Northern Africa Hypothesis would predict that the Gitanos in Spain are comprised of both Roma (from the Northern and Mediterranean routes) and Dom (from the Northern African route).
- This, in turn predicts that the Gitano language should contain elements of both Romani and Domari; or, at least, significant evidence of a Domari substratum.

Caló

• The language of the Gitanos is called *Caló*. Bakker 1995 classifies it as a *Para-Romani* language – that is, it consists of a specialized vocabulary embedded in Spanish, utilizing Spanish phonology, morphology, and syntax.

(3)	<i>kamel-</i> 'love' (< Skt. <i>kama</i>):		Sg.	Pl.
		1 st	kamelo	kamelamos
		2^{nd}	kamelas	kamelais
		3^{rd}	kamela	kamelan

- The Gitanos' language underwent considerable attrition, perhaps because of forced assimilation efforts.
- Bakker (1995:142) notes that evidence of Spanish morphology is evident in the first recorded example of Caló, (17th century).
- Caló sentences, collected in 1818, also contains considerable Spanish:

(4)	a.	Ochanaba mangue lo que chil	е.
		/	•

- 'I know not what you tell me.'b. Gill*ate de mi que no te pueda in*dic*ar*.'Get out of my sight.'
- McLane (1987), based on anthropological work in the Gitano community of Gaudix, Granada, estimates that fluent speakers actually use no more than 100 or so Caló words in their speech
- Larger vocabularies were compiled in the 19th and early 20th centuries (e.g. Borrow 1843, Quindalé 1870, Rebolledo 1909).
- As the name *Para-Romani* suggests, the assumption has been that Caló consists of Romani vocabulary embedded in a Spanish matrix language.
- However, the presence of a Domari element in has not been investigated.
- Bakker (1995:132) notes that a distinguishing characteristic of Caló is a large vocabulary not shared with other varieties of Romani. If some of this turned out to be Domari, it would lend support to the Northern Africa Hypothesis.

Caló: Para-Romani or Para-Romani/Domari?

- In order to determine whether Caló has a Domari substratum, it was compared to both Romani and • Domari along three dimensions: cognates, loanwords, and sound correspondences.
- 197 of the 200+ lexical items in the Manchester Romani Project wordlist were searched for cognates between Romani and Caló, Domari and Caló, or between all three languages.
- The Romani forms are the Kaldaraš forms from the Manchester database. Occasionally, data from • other European Romani dialects are brought in to clarify a cognate relation.
- The Domari data come from Macalister's 1914 dictionary of the Jerusalem dialect. •
- The Caló data are primarily from Rebolledo (1909), but occasionally from Quindalé (1870). •
- A few forms are from current usage, as reported in Ropero Núñez (1978), McLane (1987), Leigh (1998), or elicited; these are underscored.
- Note Caló has adopted the phonology of (Andalucian) Spanish.
- Also there are regular morphological differences between Romani and Caló. For example, Caló nouns and adjectives typically end in accented $-\dot{o}$ (masc.) or accented -i.
- Citation forms for verbs end in -el in Romani (based on third person singular present), -ar (third • person, singular, progressive) in Domari, and -ar in Caló.
- The identity in the last two cases is accidental, as the -ar suffix in Caló is the Spanish infinitival suffix. •

a.	<i>Kaldaraš Romani</i> guruvni	Domari górwi	<i>Caló</i> burí	'cow'
b.	mak ^h	makíli	mačá	'fly'
c.	bakri	bákri ('ewe')	brakí	'sheep
d.	pəř	pe:t	poria ('womb')	'belly'
e.	čuči	čič	čučaj, <u>čučá</u>	'breast'
f.	kan	kan	kan	'ear'
g.	jak ^h	íki	akí, <u>sakaj</u> ¹	'eye'
h.	bal	wal	<u>bal</u> , <u>bales</u>	'hair'
i.	vast	xast	baste ² , <u>baes</u>	'hand'
j.	koř	gúrgi	korrajá ('necklace')	'throat'
k.	šib ³	ĵib	čipí	'tongue'
1.	manřo	móna	<u>manró</u>	'bread'
m.	anřo	ána	anró	'egg'
n.	anav	nam	naw	'name'
0.	des	di:s	čibel ⁴	'day'

(5) Possible cognates in all three languages (Romani, Domari, and Caló)

¹ The initial /s/ in Caló /sakaj/ comes from a reanalysis of the final /s/ of the plural definite article /los akajs/

> /lo sakajs/. The final /s/ is lost as part of a regular sound change in Andalucian Spanish, hence, /(s)akajs/

> /(s)akaj/. ² /v/ > /b/ / #____ is a regular sound change in Spanish. ³ Cf. Polska Roma $\ddot{c}^h ib$.

⁴ Cf. Lešaki Polska Roma /dives/; Caló /č/ may be due to due to palatalization.

a. b. c. d.	<i>Kaldaraš Romani</i> čořes , nasul dilo kir k ^h ajni	<i>Domari</i> (čámda) (lála 'dumb') (mórže) (čmári)	<i>Caló</i> <u>čorró</u> , nasaló ('sick') dililó, <u>lililo, lileá</u> kiria kañí	'bad' 'stupid' 'ant' 'chicken'
е.	žukəl	(snóta)	<u>čukel, čuké</u>	'dog'
с. f.	pišom	(kéči)	paxuma ($\tilde{s} > x$)	'flea'
g.	papin	(tílla-čmári)	papín	'goose'
h.	grast	(jégir)	grasté, <u>gras</u>	'horse'
i.	balo	(bug, xúgi)	baličé, <u>baličó</u> ('pork')	ʻpig'
j.	šošoj		xoxoj $(\breve{s} > x)$	'rabbit'
k.	ruv		orú	'wolf'
1.	naj	(úngli)	naj ('fingernail')	'finger'
m.	čang	(dó:ni)	čanklí	'knee'
n.	punřo	(paw)	pinré ('foot') <u>pinrel</u>	'leg'
0.	muj	(zári)	<u>muj</u> , <u>miu</u>	'mouth'
p.	abdin	(gúlda)	agín	'honey'
q.	ambrol		bronda	'pear'
r.	del	(xúĵa)	<u>undebel</u> , <u>debel</u>	'God'
S.	angar	(wášri)	angar	'coal'
t.	udar	(kápi)	burdó	'door'
u.	sumnakaj	(zerd)	<u>sonakaj</u>	'gold'
V.	čjar	(gas)	ča	'grass'
W.	pativ		pačiba	'honor'
Х.	baxt	(mirité:k, rizq)	baxí ('fortune')	'luck'

(6) Possible cognates between Romani and Caló (Domari non-cognate in parenthesis)

86 out of 197 possible cognates between Romani and Caló.⁵ Out of these 27 were also cognate in • Domari, but this is may be because both Romani and Domari are Indic languages.

Are there any examples shared by Domari and Caló that are not also cognate in Romani? •

(7) shows, there are a few cases where Domari and Caló are cognate and the Kaldaraš form is not. • However, once we look at other Romani dialects, we find no clear cognates that are exclusive between Domari and Caló:

(7) Possible cognates between Domari and Caló (Kaldaraš non-cognate in parenthesis)

	Kaldaraš Romani	Domari	Caló	
a.	(unto)	gir	kir ⁶	'butter'
b.	(tiral)	kir	kirá ⁷	'cheese'
c.	(div)	gésu: ('corn')	gi ('wheat') ⁸	'wheat/rye'
d.	(žuvli)	káĵĵi	kačí	'woman'
e.	(rudil)	mángiš-kerar	<u>mangar⁹</u> ('rob')	'beg'
f.	(šinel)	kérar	kerar ¹⁰	'do'
g.	(mudarəl)	márar	marar ¹¹	'kill'

⁵ In 12 cases there was no attested Domari form.

⁶ Cf. Lajesha/Kishinevski (Moldavia) *k^hil*; Hungarian Vend Romani *kiralin*.

⁷ Cf. Lešaki Polska Roma kirał.

⁸ Cf. Lajesha / Kishinevski giw.

⁹ Cf. Lešaki Polska Roma *mangel*.

¹⁰ Cf. Lešaki Polska Roma *kerel.*¹¹ Cf. Serbian (*u*)mar-; Romacilikanes (Greece) marava.

- With the possible exception of (5d), each of these Domari-Caló cognates find a clear cognate in another dialect of European Romani.
- The lone exception, *kači* 'woman', is clearly related to *gači* 'Non-Gypsy woman', which finds a Kaldaraš Romani cognate in *gažo* 'non-Gypsy'.
- Unless we take this one example as evidence for a Domari substratum, we find that cognates do not support the Northern Africa Hypothesis.
- Also significant is the fact that over half of the Caló words in the sample were cognate with neither Romani nor Domari. This shows that the large non-Romani vocabulary commented on by Bakker (1995) is not due to a Domari influence.

Loanwords

- Given the history of both Roma and Dom in Diaspora, their languages have acquired loanword vocabularies from the languages they have been in contact with.
- Romani shows a significant number of Iranian, Armenian, and Greek borrowings, as well as a few from Georgian.
- Domari has also has Iranian borrowings, as well as a large number of Arabic loans.
- While both Romani and Domari have Iranian elements, the overlap in their Iranian loanwords is not great (Hancock 1995).
- The tables in (6-10) show that Caló follows only the Romani pattern:
 - It exhibits a large number of Greek borrowings.
 - It has the Armenian and Georgian borrowings shared by Romani.
 - Its Iranian borrowing follow the Romani, but not the Domari pattern.

(8) Greek Loanwords in Romani (parenthesis indicate forms not borrowed)

	Romani	Domari	Caló		Greek
a.	foros	(ujára)	foró	'town'	fóros 'market'
b.	drom	(pand)	dron	'road'	drómos
c.	zumin		sumí	'soup'	zumí
d.	kokalo	(xar)	kokalé	'bone'	kókkal
e.	xoli	(ris)	xojin	'anger'	xolí
f.	karfin	(mi:x)	karfialar (v)	'nail'	karfi
g.	kurko		kurkó ('Sunday	y') 'week'	kyriakí 'Sunday'

(9) Armenian and Georgian Loanwords in Romani

	Romani	Domari	Caló		
					Armenian
a.	bov	(móna-kerna)	(sosimbo)	'oven'	bov
b.	grast	(jégir)	grasté	'horse'	grast
c.	kotor	(hítar)	kotoré	'piece	kotor
d.	pativ		pačiba	'honor'	patiw
					Georgian
e.	k ^h iliv		kijaba	'plum'	khil-avi
f.	čamčali		(sosimbre)	'eyelash'	camcami

(10) Iranian Loanwords in Romani

	Romani	Domari	Caló		
					Persian
a.	ambrol		bronda	'pear'	amru:d
b.	res-	(hád-hoč)	(bergorear)	'arrive'	ras-i:dan
c.	avgin	(gúlda)	angin	'honey'	angubi:n
d.	diz	(ujára)	(gaw)	'town'	diz
e.	baxt	(mirité:k, rizq)	baxí ('fortune')	'luck'	Pers. baxt,
					Kurd. bexit
i.	sir		sirí	'garlic'	Pers. & Kurd. sir

- These data clearly demonstrate that Caló shares a significant number of borrowings typical of Romani.
- However, in order to argue for a lack of a Domari substratum, we have to show that Caló lacks borrowings typical of Domari.
- Hancock (1995) notes that Domari has a significant Iranian vocabulary, but it is largely disjoint from the Iranian that found in Romani.
- If Caló shows a Domari substratum, we expect to find some of the Iranian borrowings unique to Domari.
- As the examples in (11) show, the Iranian loanwords that Hancock describes as unique to Domari are not found in Caló:
- (11) Iranian Loanwords in Domari

	Kaldaraš Romani	Domari	Caló		
a.	(p ^h abaj)	sib	(poba)	'apple'	Pers. sib
b.	(del)	xúĵa	(undebel)	'god'	Pers. xudá,
					Kurd. xudan
c.	(pala)	pači	(andá)	'after'	Kurd <i>paši</i>
d.	(vorzo)	ĵau	(roso)	'barley'	Pers. ĵo
e.	$(\text{sor})^{12}$	kuč	(čon)	'beard'	Pers. xu:ša
f.	(punřo)	pau	(<u>pinré)</u>	'foot'	Pers. pa

• The largest source of borrowings into Domari is Arabic. The examples in (12) are those from the sample that Maccalister (1914) identifies as Arabic borrowings.

(12) Arabic Loanwords in Domari

	Kaldaraš Romani	Domari	Caló		Arabic
a.	(čistil)	imsáħ-kerar	(alipé)	'clean'	imsah 'erase'
b.	(omida)	du:d	(kirmó)	'worm'	duda
c.	(jasfa)	dé:mi'	(bielima)	'tear'	dama '
d.	(lumina)	dáwwi	(mumelí)	'light'	daw'
e.	(pozynari)	že:b	(kisía)	'pocket'	ĵi:b

• The data in (11-12) show that there is no evidence from loanword vocabulary for a Domari substratum in Caló.

¹² Cf. Finnish Romani *čo:n*.

Sound Correspondences

- The final criterion for distinguishing Romani from Domari comes from a series of regular sound correspondences.
- Romani and Domari differ in their reflexes of a number of Sanskrit phonemes in their shared Indic vocabulary, as first discussed in Sampson (1926).
- The Northern Africa Hypothesis predicts that Caló should show evidence of a Domari substratum.
- Hence, we might expect to find that Caló sometimes shows reflexes associated with Domari and not Romani
- The tables in (13-17) show the major correspondences between Romani and Domari (Sampson 1926) and the Caló reflexes.
- It is clear that Caló consistently patterns with Romani, not with Domari.

(13) Rom. vl aspirated stops : Dom. vd stops (< Skt. voiced aspirates)

a. b. c. d. e. f. g.	<i>Romani</i> k ^h am k ^h as čib p ^h en p ^h ral p ^h enel p ^h us	Domari gam gas ĵib be:n bar bis	<i>Caló</i> ¹³ kam kas ('firewood') čibí penčí plal <u>penar</u> pus	'sun' 'hay' 'tongue' 'sister' 'brother' 'say' 'straw'	$\begin{array}{l} (\text{Skt. } g^h arma) \\ (\text{Skt. } g^h a : sa) \\ (\text{Skt. } jihva) \\ (\text{Skt. } b^h agini:) \\ (\text{Skt. } b^h ra:tr) \\ (\text{Skt. } \sqrt{b^h an}) \\ (\text{Pkr. } * b^h usa) \end{array}$	
(14)	Rom. r : Dom. ţ (< Skt. ţ)					
a. b. c.	<i>Romani</i> bar per varo	<i>Domari</i> waţ peţ ató:s	<i>Caló</i> bar poria ('womb') xaroj	'stone' 'belly' 'flour'	(Skt. <i>vaţa</i>) (Skt. <i>peţa</i>) (Hnd. <i>a:ţa:</i>)	
(15)	Rom. l : Dom. r (< Skt. t)					
a. b c. d. e.	<i>Romani</i> gili k ^h il p ^h ral ĵala	<i>Domari</i> gire gir bar ĵar	<i>Caló</i> gijabó [l > l ^y > j] kir ¹⁴ plal čala	'song' 'butter' 'brother' 'he goes'	(Skt. $gi:ta$) (Skt. $g^h rta$) (Skt. $b^h ra:tr$) (Skt. $ya:ti$)	

¹³ The Caló reflexes of Romani voiceless aspirated stops are voiceless unaspirated stops, due to the fact that Spanish lacks aspirated stops. ¹⁴ Given that the Sanskrit form has both /r/ and /t/, the Caló /r/ may be a reflex of Sanskrit /r/, with a loss of

¹⁴ Given that the Sanskrit form has both /r/ and /t/, the Caló /r/ may be a reflex of Sanskrit /r/, with a loss of /t/. Of course, this could be true of Domari as well, in which case, this example is inconclusive. Note that Hungarian Vend Romani *kiralin* retains reflexes of both the Sanskrit /r/ and /t/.

(16) **Rom. b :Dom. w (< Skt. v)**

	Romani	Domari	Caló		
a.	bal	wal	bal	'hair'	(Skt. va:la)
b.	beš	wé:sar	bexelar	'sit'	(Skt. √ viš)
c.	biš	wi:s	bin	`20 '	(Skt. vimšati)
d.	brišind	warsínda	brixindia	'rain'	(Skt. va:ri)

(17) **Rom v : Dom. m (< Skt. m)**

	Romani	Domari	Caló		
a.	nav	nam	naw	'name'	(Skt. na:man)
b.	p ^h uv		pu	'earth'	(Skt. $b^h u$:mi)

- The sample surveyed in section 3 failed to show evidence of a Domari substratum in Caló.
- No compelling case was found in cognates, loanwords, nor sound correspondences.
- This result is probably uncontroversial it reinforces what most have assumed, namely, that Caló evolved from a variety of Romani.
- It does bear on the sustainability of the Northern Africa Hypothesis, which continues to be widely cited in popular works on Spain's Gitanos.
- Nevertheless, in addition to lacking historical evidence, the Northern Africa Hypothesis lacks linguistic evidence.