

FROM IBERIAN ROMANI TO IBERIAN PARA-ROMANI VARIETIES

ZUZANA KRINKOVÁ

KAROLINUM

From Iberian Romani to Iberian Para-Romani Varieties

PhDr. Mgr. Zuzana Krinková, Ph.D.

Reviewed by:

Doc. Mgr. Ivo Buzek, Ph.D. (Olomouc, Czech Republic)

Prof. Dr. Ignasi-Xavier Adiego-Lajara (Barcelona, Spain)

Jorge M. F. Bernal (Buenos Aires, Argentina)

Published by Charles University in Prague,

Karolinum Press

Proof-read by Pearl Harris

Cover and layout by Jan Šerých

Typeset by Karolinum Press

First edition

© Charles University in Prague, 2015

© Zuzana Krinková, 2015

ISBN 978-80-246-2936-9

ISBN 978-80-246-2949-0 (online : pdf)



Charles University in Prague
Karolinum Press 2015

www.karolinum.cz
ebooks@karolinum.cz

TABLE OF CONTENTS

INTRODUCTION ---- 10

1 SEVERAL NOTES ON ROMANI ---- 13

- 1.1 Proto-Romani, Early Romani, Common Romani ---- 14
- 1.2 Classification of Romani Dialects ---- 15
- 1.3 Para-Romani Varieties ---- 21
 - 1.3.1 Delimitation of Para-Romani Varieties ---- 21
 - 1.3.2 Language Structure of Para-Romani Varieties ---- 22
 - 1.3.3 Genesis of Para-Romani Varieties ---- 23
 - 1.3.4 Examples of Para-Romani Varieties ---- 25

2 ROMANI LANGUAGE ON THE IBERIAN PENINSULA ---- 27

- 2.1 The History of Romani Language in the Iberian Peninsula ---- 28
- 2.2 Classification of the Iberian Romani Dialect and Diversification of Iberian Para-Romani Varieties ---- 34
- 2.3 Present Situation of Iberian Para-Romani Varieties ---- 38
 - 2.3.1 Caló ---- 38
 - 2.3.2 Basque Para-Romani ---- 40
 - 2.3.3 Catalan Para-Romani ---- 40

3 SOURCES NEEDED WHEN STUDYING CALÓ AND OTHER IBERIAN (PARA-)ROMANI VARIETIES ---- 41

- 3.1 Introduction ---- 42
- 3.2 The Survey of Sources ---- 44
 - 3.2.1 The Sources Needed when Studying Caló ---- 44
 - 3.2.2 Sources Needed when Studying Catalan (Para-)Romani ---- 52
 - 3.2.3 The Sources Needed when Studying Basque Para-Romani ---- 52
 - 3.2.4 The Sources Needed when Studying Brazilian Para-Romani ---- 53

4 LINGUISTIC DESCRIPTION OF IBERIAN PARA-ROMANI VARIETIES ---- 55

- 4.1 Several Comments on Spelling and Phonetic Transcription ---- 56
 - 4.1.1 Spelling Peculiarities of Spanish Caló ---- 58
 - 4.1.2 Spelling Peculiarities of Catalan (Para-)Romani ---- 60
 - 4.1.3 Spelling Peculiarities of Basque (Para-)Romani ---- 63
- 4.2 Phonology ---- 63
 - 4.2.1 Vowels ---- 63

4.2.2	Glides and Diphthongs ----	64
4.2.3	Consonants ----	65
4.2.4	Sonorants ----	85
4.2.5	Consonant Clusters ----	87
4.3	Sporadic Sound Changes ----	89
4.3.1	Prothesis and Apheresis ----	90
4.3.2	Epenthesis ----	93
4.3.3	Epithesis ----	96
4.3.4	Syncope ----	97
4.3.5	Apocope of Consonants and Syllables ----	98
4.3.6	Vowel Shifts and Alternations ----	99
4.3.7	Consonant Alternations ----	100
4.3.8	Metathesis ----	103
4.3.9	Substitution of Consonants in Consonant Clusters and in Final Position of the Word ----	104
4.3.10	Palatalisation of /di/, /ti/, /ni/, /li/ ----	104
4.3.11	Voice Opposition ----	105
4.3.12	Gemination ----	106
4.3.13	Contamination ----	107
4.4	Reflection of Romani Morphology in Iberian (Para-)Romani Varieties ----	107
4.4.1	Nominal Morphology ----	107
4.4.2	Verbal Morphology ----	140
4.5	Several Notes on Syntax ----	159
4.5.1	The Noun Phrase ----	160
4.5.2	The Verb Phrase ----	162
4.5.3	Possessive Construction ----	162
4.5.4	Negation ----	162
4.6	Vocabulary ----	163
4.6.1	General Issues of Romani Vocabulary ----	163
4.6.2	The Issue of Iberian Para-Romani Vocabulary ----	164
4.6.3	Semantic Domains of Inherited Lexicon and their Representation in Iberian (Para-)Romani Varieties ----	165
4.6.4	Preserved Loanwords in Iberian (Para-)Romani Varieties ----	172

Conclusion ---- 185

References ---- 191

Resumen ---- 199

Index of Languages and Linguistic Terms ---- 205

Appendix: Survey of Romani Etymologies in Iberian (Para-)Romani Varieties ---- 209

LIST OF TABLES AND FIGURES

Table 1.	Models of classification of Romani dialects ----	16
Figure 1.	North-South division in Romani dialects ----	20
Figure 2.	Archaic forms of demonstratives in Romani dialects ----	21
Table 2.	Survey of transcription of chosen sounds ----	57
Table 3.	Graphic transcription of chosen phonemes in Romani, Spanish, Catalan and Basque ----	57
Table 4.	Graphic recording of problematic consonants in Catalan (Para-)Romani ----	62
Table 5.	Graphic recording of problematic consonants in Basque (Para-)Romani ----	63
Table 6.	Reconstructed phonemes in Early Romani ----	65
Table 7.	Phonemes in modern Spanish ----	66
Table 8.	Phonemes in medieval Spanish (beginning of the 15 th century) ----	66
Table 9.	Phonemes in modern Catalan ----	67
Table 10.	Phonemes in modern Basque ----	67
Table 11.	Reconstructed nominal stems in Early Romani ----	108
Table 12.	Agglutinative case markers ----	113
Table 13.	Adjective inflection ----	116
Table 14.	Reconstructed forms of definite article in Early Romani ----	118
Table 15.	Reconstructed deictic and anaphoric expressions in Early Romani ----	120
Table 16.	Reconstructed forms of personal and possessive pronouns in Early Romani ----	122
Table 17.	The forms of personal and possessive pronouns in Common Romani ----	122
Table 18.	Reconstructed interrogative pronouns and adverbs in Early Romani ----	127
Table 19.	Reconstructed indefinite pronouns in Early Romani ----	128
Table 20.	Romani numerals documented in Iberian (Para-)Romani varieties ----	130
Table 21.	Supposed stem types and their perfective markers in Early Romani ----	142
Table 22.	Paradigm of Romani copula ----	144
Table 23.	Supposed person concord markers in Early Romani ----	147
Table 24.	Reconstructed adaptation of loan verbs in Early Romani ----	157

LIST OF ABBREVIATIONS

ABL	ablative	Lom.	Lomavren
ACCUS	accusative	M	masculine
ADJ	adjective	Maced.	Macedonian
Arab.	Arabic	NOM	nominative
Balarus.	Belarussian	OBL	oblique
Basq	Basque (Para-)Romani	OIA	Old Indo-Aryan
Basq.	Basque	PART	participle
Bulg.	Bulgarian	Per.	Persian
C	consonant	PL	plural
Cat	Catalan (Para-)Romani	Pol.	Polish
Catal.	Catalan	Port.	Portuguese
Cr.	Croatian	Rom.	Romani
Cz.	Czech	Roman.	Romanian
DAT	dative	Rus.	Russian
Domar.	Domari	Serb.	Serbian
F	feminine	SG	singular
Fr.	French	Sk.	Slovak
GEN	genitive	Sl.	Slovenian
Germ.	German	Slav.	Slavic
Gr.	Greek	Span.	Spanish
Hind.	Hindi	SUBST	noun
Hung.	Hungarian	Tur.	Turkish
Ind.	Indian	Ukr.	Ukrainian
INSTR	instrumental	V	vowel
LOC	locative		

Majarados sinareis, pur sangue aborrecieren os manuces, y sangue buchararen de junos, y sangue curararen, y chibaren abrí o nao de sangue, como choro, por o Chaboro e manu.

Embéo e Majaró Lucas, 6:22

Blessed are ye, when men shall hate you, and when they shall separate you from their company, and shall reproach you, and cast out your name as evil, for the Son of man's sake.

Gospel according to St. Luke, 6:22

INTRODUCTION

The incentive to write this work came quite by accident several years ago in Valencia when I was scrolling through a dictionary of Spanish argot, in which several words strikingly resembled colloquial expressions I knew from Hungarian. This similarity intrigued me and two years later a sociolinguistically oriented thesis emerged, dealing with a comparison of the occurrence of words of Romani origin in contemporary colloquial Spanish and Hungarian. Of course the work required me to become familiar with at least the basics of the Romani history and language, which I did in the form of self-study. The issue began to intrigue me so much that I decided to continue the idea and pursue a deeper study of the Iberian Romani language, especially Spanish Caló and its influence on the Spanish language. This resulted in the PhD thesis entitled *Mutual contact of Romani, Spanish and other languages of the Iberian Peninsula* (Krinková 2013b), on which this book is largely based.

I am fully aware that my interest in Iberian Romani, Caló and language contact is far from ground-breaking. This work builds on a number of scientific publications, both from Spanish (C. Clavería, currently I.-X. Adiego and others) and also from leading European contemporary linguists dealing with Romani, such as N. Boretzky and P. Bakker. In the Czech Republic, the issue of *Gitanisms* and dictionaries of Caló are dealt with by I. Buzek (e.g. *La imagen del gitano en la lexicografía española*, 2010). An overview of the available resources on Iberian Para-Romani varieties is provided in a separate chapter.

The works to which I refer are mainly articles or partial studies only dealing with selected issues of the relevant theme. However, unlike the aforementioned works, this book provides the first systematic and comprehensive processing of the grammar and vocabulary of Iberian Romani and Para-Romani varieties.

At the forefront of my interest are varieties of Romani that developed in the Iberian Peninsula after the arrival of the Roma in the 15th century. To describe these variants, I have used extensive linguistic material (in particular, dictionaries and secondary sources on Iberian Para-Romani), from which I was able to extract a large amount of Romani etymology. Due to my Hispanic qualifications, I focus in particular

on Spanish Caló; however, I also deal with other varieties, mainly Catalan and Basque Para-Romani. At times I also mention Brazilian Para-Romani, but do not go into too much detail about it in this work, preferring to refer interested readers to other literature. On the contrary, I pay great attention to a variant documented in Portugal which clearly derives from Southern Spanish Caló. Quite apart from my interest, there are the inflectional Romani dialects, which arrived in Spain with the more recent waves of Roma immigration during the 19th and 20th centuries.

The first chapter aims to briefly introduce the reader to the issues of the Romani language and Romani linguistics. Unless otherwise stated, I refer here mainly to the introduction to Romani linguistics given by Matras (2002). I clarify certain terms later used (e.g. the term ‘Para-Romani’), and point out the problem areas of contemporary Romani studies which are crucial to this work (e.g. reconstruction of Early Romani, classification of Romani dialects).

My primary hypothesis is that the Roma people brought the Romani dialect to the Iberian Peninsula in the 15th century. This dialect, from the current point of view, is quite conservative, and we can assume that in many respects it was not very different from the (reconstructed, undocumented) phase of so-called ‘Early Romani’. Subsequently, I look at phonological, morphological, syntactic and lexical developments from this archaic inflectional Iberian Romani to the Para-Romani varieties.

In the chapter on Phonology, I characterise the phonetic development of all three of the aforementioned forms of Iberian Para-Romani varieties. I also focus on some as yet unknown or only partially described phenomena (e.g. the development of sibilants and nasalisation). The chapter on Phonology also includes the issue of spelling in Iberian Para-Romani, whose peculiarities can often lead to misinterpretation of the information contained in source material. I also place emphasis on the contact with Spanish and other languages and language variants of the Iberian Peninsula. This language contact has been occurring since the 15th century; for this reason, I take into account not only the current condition of contact languages but also their diachronic evolution, which is particularly important for the phonological subsystem of Iberian Para-Romani varieties.

The chapter devoted to a description of the remnants of the Romani morphological subsystem is quite extensive, due in particular to the fact that Romani morphology is described only very marginally or not at all in the works of Iberian Para-Romani, because for the most part it is no longer productive. In my opinion, however, lexicalised remnants of archaic Romani morphology provide very valuable information, not only for the reconstruction of the inflectional Iberian dialect, but also for the reconstruction of the development of Romani as a whole.

In the chapter on Vocabulary, I deal with the Indian vocabulary and pre-European loanwords, I also pay particular attention to loanwords from the Greek and Slavic languages.

At this point, I would like to thank the people without whom this work would not have been possible, or at least not in its current form. I would firstly like to mention my colleagues from the Faculty of Arts, Charles University, Associate Professor of

Hispanics, Dr Petr Čermák, PhD, Professor Dr Bohumil Zavadil, CSc for their longtime support during my studies and Dr Viktor Elšík, PhD, expert on Romani, for his precious advice. My other thanks belong to the reviewers: Dr Ivo Buzek, PhD, Associate Professor of Hispanics from the Faculty of Arts, Masaryk University in Brno, Dr Ignasi-Xavier Adiego, Full Professor of Indo-European Linguistics from the University of Barcelona and José M. F. Bernal, President of AICRA (Asociación Identidad Cultural Romaní de Argentina). Further I thank Pearl Harris for the revision of the English text. I am also grateful to my husband, Ondřej Krinke, especially for his patience during the creation of this work. I also want to thank my parents, Helena and Michal Čenger, and my sister, Helena Charles, for their long-term support and assistance in looking after my young son.

1. SEVERAL NOTES ON ROMANI

1.1 PROTO-ROMANI, EARLY ROMANI, COMMON ROMANI

Contemporary Romani dialects contain a series of conservative and innovative features due to which Romani differs from other modern Indo-Aryan languages, including Indian languages in the diaspora. The entirety of these development features is reflected in the first development phase of Romani as an independent language, which is known as Proto-Romani (cf. Matras 2002; Elšík 2006). Proto-Romani dates to the period when it distinctly diversified itself from other related languages. Nevertheless it is difficult to establish exactly when this happened since no written documentation of this phase has been discovered so far. When reconstructing it is necessary to make use of a comparison of related words of old Indo-Aryan languages and their modern Indian successors in the region of India and in the diaspora with present-day Romani dialects. Romani shares a part of language changes with the other languages in the territory of India; some changes are shared by Romani and Indian languages in the diaspora (e.g. Domari or Lomavren) and other changes are typical only for Romani.

As an example of the reconstruction of the Proto-Romani form there is an oblique case of the demonstrative SG.M **otas* > *oles*, SG.F **ota* > *ola*, even though the forms *oles* and *ola* have been preserved only in a few Romani dialects. The reconstructed forms can be however supported by other proofs: 1) they appear in a more recent form as *od-oles*, *od-ola*, 2) they survive in the contracted form *les*, *la* in oblique case of the pronoun of 3SG, 3) they correspond with the Domari demonstratives SG.M *oras*, SG.F *ora* and 4) the old Indian demonstrative stem *t-* is well attested and the change of the old Indian /t/ > /l/ (/t/ > /r/ in Domari) is regular.

Another phase, and much better documentable, is Early Romani (cf. Matras 2002; Elšík 2006). It is characteristic due to its adoption of productive Greek morphology (called athematic or xenoclitic morphology) applied mainly to loanwords and other structural innovations drawing from contact with Greek, such as the emergence of the proposed definite article.¹ Early Romani is not documented in the written form;

1 Fraser (1998) states that in the Greek speaking territory some significant phonetic changes occurred: stem *m* turned to *v* (Sanskrit *nāman* > *nav*), initial and stem *h* turned to *j* or *v* (Sanskrit. *hāsta* > *vast*). Romani was also enriched by means of the phoneme *f* in Greek loanwords (such as *karfín*).

however its birth dates back to the Byzantine period of around the 10th or 11th century. The period of Early Romani ends with a rise in the present dialects and their dispersal in Europe and it is dated on the basis of hints in historical sources to the 14th century. The Early Romani forms are conservative structures that have survived so far only in some dialects.

A good example of an Early Romani structure is a set of demonstratives *adava/akava*. These forms are recorded both in the most western Romani dialect in Wales and in one of the most eastern dialects, Southern Balkan Arli (and, as I state further on, also in Iberian Romani). In other contemporary dialects we may find simplified and reduced forms such as *dava/kava* or *ada/aka* or innovative forms such as *kado/kako*.

In Early Romani we may in phonology assume a phoneme /ř/ (e.g. in the word *řom* 'Rom') the phonetic quality of which is unknown. It could also be the uvular/R/ which has survived so far e.g. in Kelderaš Romani or the Proto-Romani retroflex /ḍ/ > /*], *ř/ (cf. Indo-Aryan *ḍom*). In many Romani dialects then this /ř/ has merged with /r/.

One of the most important tasks that contemporary comparative Romani dialectology has to face is to state which elements from present-day Romani dialects can be dated to the period of Early Romani or even Proto-Romani. On the other hand it may seem that many forms and structures have been carried over from the Early Romani period in an almost unchanged form, since they are shared by most of the dialects. In this work, I shall refer to these forms as representing Common Romani (cf. Matras 2002). It is, for example, the numeral *oxtó* 'eight' (from Greek *oxtó*) which is only in a few dialects changed to *ofto*.

1.2 CLASSIFICATION OF ROMANI DIALECTS

The problem of classification of Romani dialects is considerably complex. Members of the Roma ethnic groups live not only in different parts of Europe, but also in the Near East, Central Asia, America etc. The Roma settlement in Europe is uneven and it does not make a language continuum in the right sense of the word. Whereas in some regions there is a high density of the Roma population and we may find in one state a great number of Roma groups which differ regarding language and culture (e.g. the Balkans and Central Europe), other regions are relatively homogeneous and the concentration of the Roma ethnics is lower here (e.g. Western Europe). In addition some Roma subgroups do not speak Romani.

When classifying Romani dialects it is necessary to especially take into account Roma migration, contact with surrounding languages and also contact among particular Roma subgroups. Roma migration into Europe started from the Balkans in the 14th and 15th centuries and has been in operation to a greater or lesser extent till now. The Roma population has always been in contact with the language of the surrounding

population of the region they have lived in. This contact with the majority population was then a source of lexical loanwords and structural innovations in all language levels.

It is necessary to note that many linguists dealing with Romani did not include in their classification the Iberian varieties of Romani (respectively also other Romani dialects of the peripheral areas). Iberian Romani is according to Boretzky and Igla (1991), Bakker and Matras (1997) classed into the Northern branch which is however very diversified.

As a better illustration here I present a table of the geographical classification of Romani dialects² with brief characteristics of the main branches.

Table 1. models of classification of Romani dialects

Supergroup	Group	Subgroup	Localisation and Nomenclature
Northern	British (BR)		Wales (<i>Kále</i>), †England, Scotland (<i>Romaničela</i>)
	North Western (NW)	Scandinavian	Finland, Sweden (<i>Kaale</i>); †Estonia, †Denmark, †Norway, †Sweden
		Sinti	Germany, Austria, Czechia, Italy etc. (<i>Cinti</i> , <i>Sinti</i>); France (<i>Manuš</i>)
	North Eastern (NE)	Western	Poland
		Eastern	Lithuania, Latvia, Estonia, Ukraine: Podolia, Russia
Iberian		†Spain, †Portugal, †Brazil (<i>Kale</i>); Basque Country (<i>Errumantxela</i>)	
Central	North Central (NC)	Western	†Czechia, West, Slovakia
		Eastern	Central and Eastern Slovakia, Southern Poland, Western Ukraine, Transylvania
	South Central (SC)	Northern	Southern Slovakia, Northern Hungary
		Vend	SW Hungary, Eastern Austria, NW Slovenia

2 The table makes use of a handout for the course of V. Elšík *Romské dialektky: dialektologie*. Accessible at://ling.ff.cuni.cz/lingvistika/elsik/ho/DRo2_Handout.pdf.

Supergroup	Group	Subgroup	Localisation and Nomenclature	
Balkan	Slovenian (SL)		Slovenia, Italy	
	Apennines (AP)		S. Italy	
	South Balkan (SB)	Northern Arli		Serbia, Kosovo: Prizren, N. Macedonia: Skopje, Kumanovo (<i>Arlija</i>)
		Southern Arli		S. Macedonia: Prilep, N. Greece: Florina, Kardica (<i>Arlija</i>)
		Sepeči		Greece: Volos, Turkey: Izmir (<i>Sepečides</i>)
		Epiros		Greece: Epiros (<i>Romacila</i>)
		Erli		Bulgaria: Cerovo (<i>Cocomaña</i>), Sofia (<i>Erlides</i>), Velingrad (<i>Yerlides</i>), Varna (<i>Bugurdžides</i>) etc.
		Ponti		Romania, Moldavia (<i>Ursara</i>), Krym, S. Russia, Georgia (<i>Kirimitika</i>)
		Iranian		N. Iran (<i>Zargari</i>)
	North Balkan (NB)	Western		Kosovo (<i>Bugurdžides</i>), Macedonia (<i>Kovača</i>)
		Kalajdži		Bulgaria: Vidin, Montana, Pazardžik (<i>Kalajdžides</i>), Romania (<i>Spoitori</i>)
		Central		Bulgaria: Sliven (<i>Nange</i>)
		Drindara		Bulgaria: Sliven (<i>Muzikantska</i>), Šumen (<i>Drindara</i>)
		Xoraxane		Bulgaria: Kaspičan (<i>Xoraxane</i>), Varna (<i>Gadžikane</i>)
	Vlax	Northern Vlax (NVL)	Lovari	Transylvania, Hungary (<i>Lovara</i>), Slovakia, Czechia, Austria, Poland, Norway etc. (<i>Čurara</i> , <i>Kherara</i> etc.)
Kelderaš			Romania, Serbia, Bulgaria, Russia, Sweden, France, America etc.	
Southern Vlax (SVL)		Northern	Romania (<i>Rakarenge</i>), Vojvodina (<i>Rabešte</i>)	
		Gurbet	Yugoslavia: Srem, Bačka, Bosna, Srbsko, Kosovo (<i>Gurbeti</i>), Monte Negro (<i>Dasikane</i>), Italy (<i>Xoraxane</i>), Macedonia (<i>Džambaza</i>), Albania	
		Southern	Bulgaria: Velingrad (<i>Rešitare</i>), Greece (<i>Filibidžija</i> , <i>Kalpazea</i> etc.)	
		Eastern	Bulgaria: Lom, Vidin (<i>Cocomaña</i>), Sindel (<i>Kalburdžudes</i>), Varna (<i>Kalajdžides</i>), Turecko (<i>Laxi</i>) atd.	
Ukrainian (UK)			Eastern Ukraine (<i>Servi</i> atd.)	
Cerhara (CE)		Transylvania, Hungary (<i>Cerhara</i> , <i>Gurvara</i>)		

For the dialects of the Balkan branch a strong Greek influence is evident which survived much longer than in the dialects that left the Balkan; further on there is also the Turkish influence. Many speakers of Balkan dialects are Muslims and many have an active command of Turkish. Dialects can be found among others also in the region of Greece, Serbia, Macedonia, Albania, Kosovo and Bulgaria. The speakers of the dialects of the Balkan branch, especially *Arli* can be found also in Western Europe where they emigrated between the 1960s and 1990s.

Vlax (Olah) branch is probably the most prominent of the Romani dialects, if we take into account the number of speakers, geographical classification and the vast documentation. This group was probably born in the Romanian speaking region. Vlax dialects do share a strong influence of Romanian on the vocabulary, phonology and adopted morphology and a number of internal innovations. During the course of history there have been several migration waves of Vlax Roma people from Romanian principalities; the most significant one is connected with the abolishment of serfdom in Romania and it lasted till the second half of the 19th century. Some dialects have been strongly influenced by Hungarian (*Lovara*). Vlax dialects can be found in many parts of Europe, especially in the region of Romania, Hungary, Bulgaria, the former Yugoslavia and Turkey. Northern The Vlax group is spread throughout Western Europe and also throughout Central Europe.

The Central branch form the dialects found in the Central European region. The Northern Central one is East Slovakian Romani, which is at present the most widespread Romani dialect in Czech Republic. Southern Central dialects embody a strong influence of Hungarian.

Other dialects are usually classed within the Northern branch. The North Western group are the Sinti-Manuš dialects which were probably born in the German speaking territory and which show a strong influence of German and share a lot of innovations. These dialects are related to dialects in Scandinavia. The North Eastern group are the dialects in the region of Poland, the Baltic region and northern Russia that also makes a coherent dialectal group. Relatively isolated are the dialects of the British and Iberian group that have become largely extinct and their remains have survived in the form of a special lexicon.

In the Northern branch there are many archaisms and some innovations to be found, some of which we may also find in the Iberian varieties of Romani. Specific sociolinguistic strategies are typical for the Northern branch. There is a substitution of the ethnic term Roma with another ethnonym (*Kale, Manuš, Sinti, Romaničal*), making of special Romani toponyms, use of nominalised genitives when creating new words (as an alternative for loanwords) and, also, the fact that the dialects of this branch are often replaced with Para-Romani varieties (see further). These features can be explained by means of a social and geographical isolation of groups and their dependence upon Romani as a secret language.

In this work I follow a relatively recent classification of dialects from a geographical-historical perspective as described by Matras (2005). The classification is based upon a premise that the borders in between particular dialects are not absolute – based

strictly upon genetic criteria in the form of a historical migration of individual sub-categories – but relative ones. Some dialects do share more common features and are thus much closer to each other than others. The structural features that differentiate the dialects are also a result of the process of changes and innovations which spread from one community to another. The results of these changes can then be marked on the map by means of isoglosses. The classification thus also takes into account, apart from the migration, a mutual contact with neighbouring Roma groups. Romani dialects, then, form a specific language *continuum* that reflects a historical spreading of structural innovations on the one hand, and the preservation of archaisms in time and space on the other.

The issue of assessing Romani innovative and conservative features is quite complex. Romani linguistics do dispose of numerous recorded language forms dating back to the Old or Middle Indo-Aryan period. A form of Early Romani can only be reconstructed on the ground of a careful comparison of Romani dialects. Let us state the following example: in Early Romani we may assume forms **andřó* ‘egg’ (< Old Indo-Aryan **āṇḍa-*) and **ařó* ‘flour’ (< Old Indo-Aryan **atṭa*). In particular dialects these two words appear in various forms. As for the historical group /ṇḍ/ we assume in Early Romani a development to */*ndř/* which in some dialects appears as /ndř/, /ndr/, /nd/, /nř/, /nl/, /rn/, /ř/ etc.; the phoneme /ř/ alternates in dialects sometimes with /r/. In some dialects in the region of the Balkans and in some peripheral dialects (e.g. in Basque Romani) the groups /ndř/ or /ndr/ have remained preserved; it is thus a conservative feature. Before an initial *a-* there may appear in some dialects the prothetic *j*-³ or *v*-⁴.

If we classify the dialects on the grounds of structural innovations, it is necessary to set which features should be incorporated into the classification. Contemporary Romani linguistics makes use of a choice of the following features when classifying the Romani dialects:

- 1) inserting of prothetic consonants: *j-aver*, *v-aver* ‘another, second’, *j-ařo*, *v-ařo* ‘flour’;
- 2) jotation and palatalisation: *kerdjom* > *kerđom* > *kerdžom* ‘I did’;
- 3) substitution /s/ > /h/: *kerasa* > *keraha* ‘we do’, *lesa* > *leha* ‘with him’;
- 4) loss of the final -s: *dives* > *dive* ‘day’, *kerdas* > *kerda* ‘he did’;
- 5) palatalisation of the consonant before i: *dives* > *džive(s)*, *džes*, *zis* ‘day’, *tikno* > *cikno* ‘small’;
- 6) palatalisation of the consonant before e: *kher* > *čher* ‘house’;
- 7) presence of a prothetic vowel: *bijav* > *abijav* ‘wedding’, *nav* > *anav* ‘name’;
- 8) simplification of the cluster */*ndř/*: *mandro* > *manro* > *maro* ‘bread’;
- 9) simplification/modification of the form of demonstratives: *akava* > *kava* > *ka-kava* > *kako*, *adava* > *ada*, *dava* > *ka-dava* > *kada* > *kado*;

3 Protetic *j-* is a result of the jotation that is a language innovation.

4 Protetic *v-* was probably born as early as in the Early Romani as a pronunciation variety when connecting the noun with the definite article: **ov-ařó*.

It is probable that the language of particular Roma clans departing from the Balkans at the end of the 14th or beginning of the 15th century was more or less uniform (although some differentiation may have occurred as early as in the Balkans). The documentation of Romani from the period up until the 17th century is very scarce. Romani documented in the sources of the 18th century nevertheless embodies largely dialectal characteristics corresponding to the present situation. We may thus suppose that the main differentiation of dialects occurred during the 16th and 17th centuries. Nomad Roma people migrated then mainly within a particular restricted territory and did not set off for any long and distanced journeys. They were thus very much influenced by the neighbouring major population, be it culturally, religiously and linguistically. It seems that during this historical period a contact between Roma groups in the region of the Habsburg monarchy and the Ottoman Empire was totally cut short which corresponds with the so-called North-South language division⁵ reflecting more features as is shown in the following map⁶.

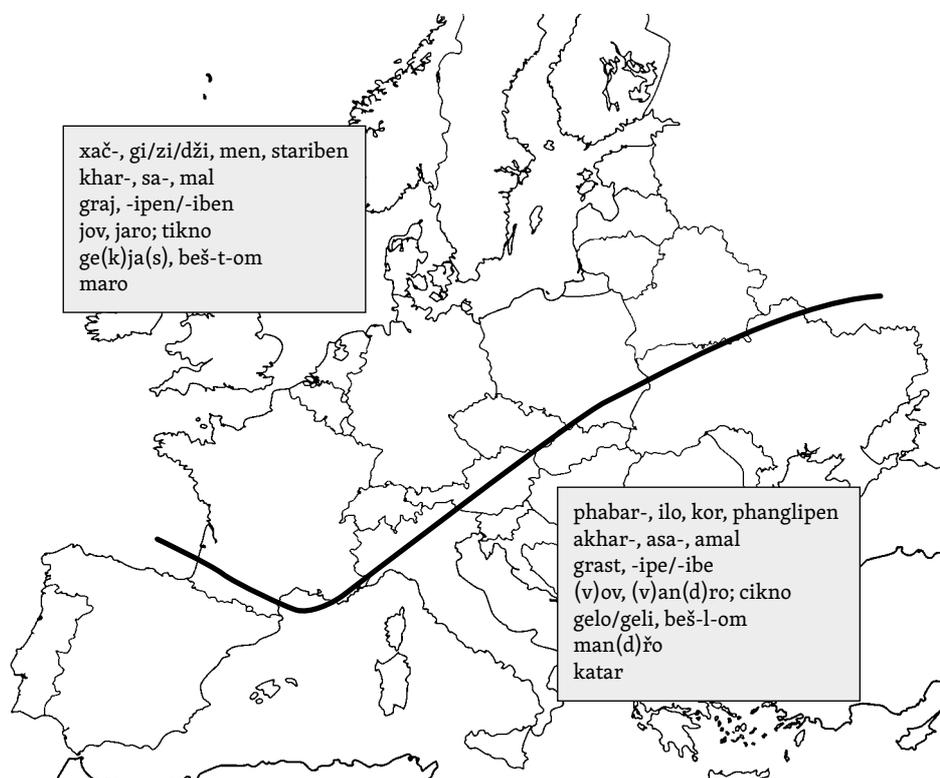


Figure 1. North-South division in Romani dialects.

5 North-South division, cf. Romani Project Manchester, also Great Divide (cf. Matras 2005: 13).

6 The maps in this chapter were created according to the maps from the Romani Project Manchester, accessible at: http://romani.humanities.manchester.ac.uk/whatis/classification/dialect_spread.shtml.

Peripheral dialects generally embody rather conservative features. This is why it is probable that the local Roma population lived isolated from other Roma groups. It is also documented by means of the following map showing the preservation of conservative forms of demonstratives (*akava*, *adava*) in peripheral regions.

1.3 PARA-ROMANI VARIETIES

1.3.1 DELIMITATION OF PARA-ROMANI VARIETIES

Apart from Romani dialects there are also language varieties where there is quite a large Romani lexicon (with relicts of the Romani inflection) incorporated into the grammar of surrounding languages. In contemporary Romani linguistics usually the term Para-Romani varieties is used to refer to these varieties.

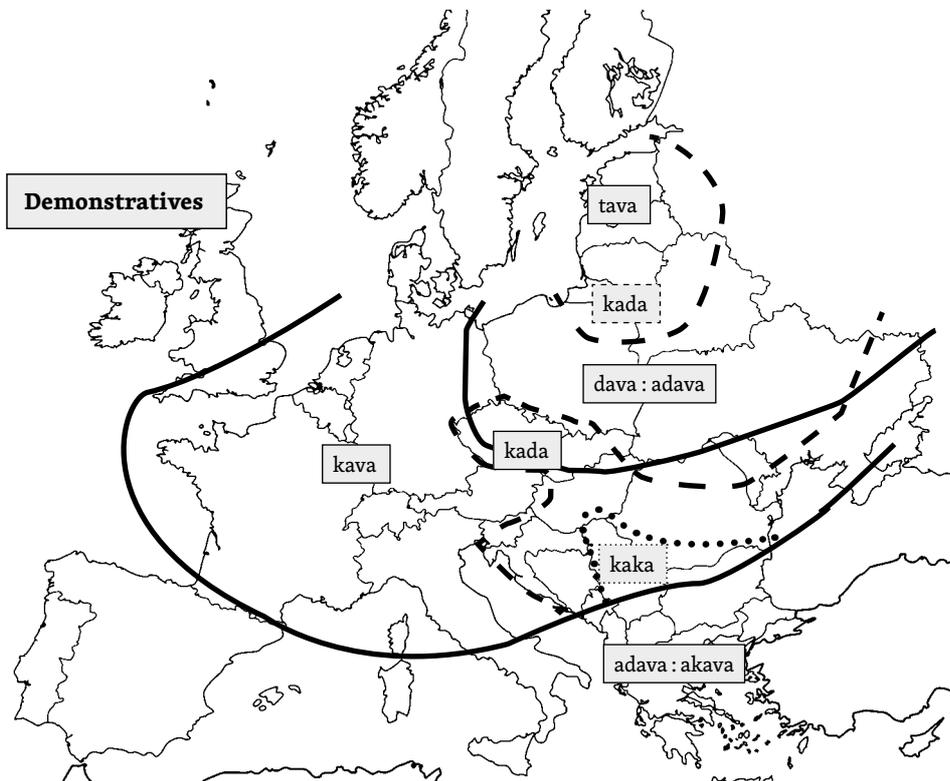


Figure 2. Archaic forms of demonstratives in Romani dialects

The term Para-Romani can be thus interpreted as a mixed language⁷, the vocabulary of which is predominantly Romani whereas the grammar (phonology, morphology and syntax) comes from the language of the majority.⁸

1.3.2 LANGUAGE STRUCTURE OF PARA-ROMANI VARIETIES

As P. Bakker and Hein van der Voort (quot. in Bakker – Cortiade 1991) state, as with all Para-Romani varieties a strict division between the grammar and lexicon can be observed. Whereas bound morphemes always come from the contact language, free grammar morphemes can originate both from the contact language and Romani, and free lexical morphemes are provided by Romani, e.g. in Caló: *no camel-o* ‘I do not want’, but also *nasti camel-o*. The words with the grammar and lexical meaning (e.g. pronouns) can be both from Romani and the contact language.⁹

It is not then a case of blending of codes within the statement (*code-switching*) or within the sentence (*code-mixing*) since the dividing line between two languages is in the word itself. Apart from that the blending of language codes presupposes a command of both language systems which are being switched in the statement. Speakers using Para-Romani varieties in most cases, nevertheless, only have a command of the grammar system of the host language and their command of the Romani lexicon is also limited. The common feature of Para-Romani varieties is also the presence of non-standard forms which are tolerated in neither of the grammar systems: such as the English Para-Romani *they pen-s* ‘they say’, *I dick-s* ‘I can see’ and *he kerav-o it* ‘he does it’ (cf. Boretzky – Iglá 1994).

-
- 7 I must stress that there is no such thing as a “pure” language and each language embodies a certain extent of interference with other languages. It is difficult to make an exact definition of the mixed language. It is usually stated that mixed languages embody differences from other cases of the influence of the language contact and the languages are mixed to such an extent that it is difficult to assess their classification from the genealogical point of view. Many linguists therefore do not class Para-Romani varieties within Romani dialects. On the other hand, as I will evidence in this work, Para-Romani varieties can sometimes provide us with a great number of valuable language materials not only lexical ones, thus enabling an insight into the already extinct Romani dialects, such as in the case of Iberian Romani.
- 8 Spanish Para-Romani Caló, similar to the argot, is a special vocabulary incorporated into the grammar structure of Spanish. This lexicon can be routinely replaced with Spanish words. Caló can also be easily related to the argot due to the fact that only a limited number of speakers have a command of it, whereas the majority does not understand its contents. In addition, a great number of European argots have drawn from Romani, which is why their lexicons can, to a certain extent, overlap. Caló differed then from the argot since it was for a long time a privilege of the *gitanos* who used it both for secret communication and as a symbol of their identity. This situation markedly changed during the 18th and 19th century due to a fashionable interest in the *flamenquismo* movement in Caló from the non-Roma population side. Some fans of the *flamenco* made dictionaries, and translated, for better authenticity, *flamenco texts* from Spanish into the Caló thanks to which the Caló became more generally familiar and ceased to be a privilege of the Spanish Roma people which can be an argument for Caló to be considered a part of Spanish.
- 9 Matras (2011) presents in his comprehensive study on language contact the following statements regarding mixed languages: (1) social factors have a greater influence upon the language contact than structural factors, (2) grammar transfer between typologically similar languages is much more probable, (3) the order of grammar transferability from the most probable to the least is as follows: the free morpheme (the most easily transferable) > aglutinating affix > fusional affix (the least easily transferable).

The language systems of Para-Romani varieties follow some general principles. The phonological subsystem has adapted to a neighbouring language, nevertheless the distribution of phonemes and their proportion in the text resembles much more Romani.

The grammar subsystem comes mainly from the majority language, even though remains of functional words and the Romani flexion appear scarcely, too. Personal pronouns are usually a mix of personal pronouns in various cases or of demonstratives. Particular Para-Romani varieties vary in a tendency of a selection of cases, e.g. English Para-Romani¹⁰ prefers the locative (*mandi* 'I, *tuti* 'you'), in Caló there are forms of the dative or instrumental (*mange*, *mansa* 'I, *tuke*, *tusa* 'you'), Scandinavian Para-Romani has chosen possessive forms (*miro* 'I, *diro* 'you'). As for other categories, Para-Romani varieties make use of Romani deixis of the place, indefinite pronouns etc. Speakers of Para-Romani varieties usually have at least some awareness of Romani nominal inflection (M, F, PL), the remains of which have been preserved. The nominal Romani suffix *-ipén* has remained partially productive, too. The forms of the verbs are usually derived from the lexical stem of 3SG (sometimes also from 1SG).

We may also find in the lexical subsystem, apart from Romani and contact elements, a tendency to make use of terms coming from argot or other sociolects of marginal population groups¹¹. The dictionaries of Para-Romani varieties usually contain around 600 Romani etymologies of the basic vocabulary. Making up new words with the help of Romani etymologies is typical for Para-Romani varieties: e.g. *sas-ti-čerikl* 'plane', lit. 'iron bird'. This can be found in Caló, too. In addition, there is also borrowing of words from local secret languages and languages of other minorities. This borrowing also shows a certain functional merging of Para-Romani varieties with sociolects of marginal communities.

1.3.3 GENESIS OF PARA-ROMANI VARIETIES

Para-Romani varieties were born independently in various places. The genesis of mixed languages has not been set unambiguously so far. The following factors play the most important role: social ones (the social status of speakers, blending of the population etc.), bilingualism, diglossia, the forgetting of one of the languages, grammar and lexical loanwords, the communication function of the mixed language (e.g. the secret language).¹²

10 A detailed description of Angloromani is provided by Matras (2010).

11 Para-Romani varieties, Caló including, were often perceived in the past as a criminal jargon, which can be evidenced by numerous definitions of the Spanish term *caló* 'jargon, criminal speech'. In Portuguese the word argot is even translated as *calão*, the etymology of which is indeed evident.

12 In linguistic circles there has been a vivid discussion on their genesis (*mixed language debate*). One of the theories supposes that Para-Romani was born due to a population blending of Roma people with local groups in the periphery. The mixed communities then created their own emblematic code. This mixed variety then replaced inflectional Romani. This process that is supposed to be rather abrupt than gradual would thus resemble a genesis of mixed languages in communities with mixed households (such as e.g. michif, cf. Bakker 1998). Boretzky (1983) stressed that preservation of the Romani lexicon is not compatible with the process of language attrition

According to the most quoted theories (cf. Matras 2011; Bakker – Van der Voort 1991) on the birth of Para-Romani varieties these mixed languages can be a result of the following processes¹³:

- (1) regrammaticalisation which is based upon a massive adopting of the grammar of the majority language (cf. Thomason – Kaufman 1988);
- (2) relexification which is based upon a massive borrowing of the lexicon from Romani to the majority language; the loss of competency in the Romani grammar comes into being later;
- (3) pidjiniisation¹⁴ and the succeeding regrammaticalisation (Romani in the position of the forbidden language could not have fulfilled all communication functions and subsequently it was reduced to such an extent that there was a necessity to start making use of the grammar of the contact language)¹⁵;
- (4) intentional creation of a new language (intertwining) which presupposes intentional mixing of Romani lexicon with the grammar of the major language with reference to the function of Romani as a secret language¹⁶. This hypothesis can be proved especially by a secret character of these mixed dialects. Their speakers are very much against revealing their speech to anybody who is not a member of the given group. In the case of English Para-Romani it even seems it is not a mother tongue but that the children are taught it as late as when they enter in contact with the majority population. Also, very often translations of such words that inflectional Romani normally adopts from local languages occur, such as expressions

(cf. Kenrick 1979) and he suggested a theory of birth as a result of the language shift towards the majority language. Matras (2002, 2011) prefers the theory of the birth of Para-Romani varieties as a result of the language shift, whereas he attributes the preservation of the Romani lexis to Romani as a symbol of identity and secret language. The discussion on the birth and character of Para-Romani varieties is not only interesting for Romani studies, but for the theoretical implication that it can also be applied in other linguistic domains, such as the pragmatics of communication, language contact, language attrition, etc.

13 For a detailed discussion see Matras (2010).

14 Matras (2011) on the contrary states that in the genesis of Para-Romani, in contrast to pidjin and creole, there is no simplification process to be found.

15 Hancock (1970) perceived mixed Romani dialects as a specific type of creoles with which they share particular features, primarily blending of two incongruous elements. On the other hand, the creole languages vary in many aspects. These were born due to European expansions, are the mother tongue, have irrespective of their location many common features, e.g. that they match the lexicon of the European language with the grammar of the local language, which is in contrast to mixed languages on the basis of Romani etc. Cortiade (1991) suggests a term *pogadi čhib* ('broken language') and replaces the process of the birth labelled formerly as creolisation with the expression *pogadisiation*. The Spanish literature dealing with Romani sometimes uses the term *pogadolecto* (cf. Jiménez González 2009). This term nevertheless evokes an idea of a decline of the language and, if not straightaway, the inevitability of its death which often does not reflect the reality. As well as in the case of the Caló Borrow speaks in the middle of the 19th century about a definitive collapse of the system that would lead to an early language death. However, even now, that is after nearly two hundred years, we may still find relicts of the Caló and we may on the grounds of particular testimonies assume that the Caló still is, at least in some communities, to a certain extent a means of the communication. Mixed language occurs sometimes apart from inflectional Romani also within one community, which is why it does not have to be a result of the language decline. This is the reason why for labelling mixed codes with Romani lexicon the term Para-Romani has started to be used, and which is at present generally acknowledged.

16 In the situation where it is forbidden to use Romani in their communication, the Roma people develop a hybrid language, the significant words of which are Romani, which is exclusive for the majority population, but which would resemble, at first, a strange dialect of the contact language.

like television, plane etc. When changing territory, it can be switched to another grammar¹⁷. Apart from the role of being a secret language Romani can be perceived also as the symbol of an ethnic identity¹⁸;

- (5) language shift which is defined by a loss of language competencies in Romani concerning the younger generations that have fully adopted the contact language, and succeeding relexification when under the influence of older generations (or a more traditionally-living community of the Roma people) some expressions from Romani were incorporated.

According to the present opinions of prominent linguists dealing with Romani, it seems probable that Para-Romani languages had been created more or less consciously with the intention to hide the content of a statement. From the perspective of Romani it was the regrammaticalisation; from the perspective of the contact language, the relexification.

1.3.4 EXAMPLES OF PARA-ROMANI VARIETIES

Para-Romani varieties have been documented mainly in West European peripheral areas where they had fully replaced inflectional Romani: English Para-Romani or *Angloromani* (*romani jib*) in Britain, Spanish Para-Romani (*Caló*) in Spain, Basque Para-Romani (*Errumantxela*) in Basque, and Scandinavian Para-Romani (*Romano*) in Scandinavia. Their speakers usually call these varieties “Romani” due to the Romani lexicon. As for other known Para-Romani varieties there are, for example, *Dortika* in Greece, *Geigelli Yürüks* in Turkey, *Romnisch* in Denmark (based on German) and others. In some cases a partial overlapping of Romani lexicon with argot occurs; some argots contain a strong Romani element. The sources usually include an enumeration of words and short sentences. The oldest documented Para-Romani is the *Caló* from the first half of the 18th century (or the end of the 17th century). Most of the sources come from the 19th century, or from the first half of the 20th century. Nowadays, Para-Romani varieties are in decline and contemporary documentation of them is rather sporadic.

Examples (Romani element in italics):

English Para-Romani (*Angloromani*)

Angloromani (1): *Mandi never dik'd a gaujo to roker Romanes.* (Matras et al. 2007: 14)

English translation: I have never seen a Gaujo (able) to talk Romanes.

17 On the grounds of a testimony of one group of Catalan Roma people it is evident that even for them the expression *Caló* means first and foremost a secret language, hidden from Gaujos, being it of any origin. When outside of Spain they deliberately use their “*Caló*” – to prevent the foreign language speaking population from understanding them.

18 As Matras (2002) states, Para-Romani can be found on web pages or gospel pamphlets, which proves this function related to the identity.

Angloromani (2): *Maw be rokkring in front of the mush and rakli!* (Matras et al. 2007: 29)

English translation: Don't talk in front of the man and [the] girl!

Spanish Para-Romani (Caló, *hispanorromaní*)

Caló (1): *Ne chiá mangue con tusa.* (Torrione 1988)

Spanish translation: No me voy contigo.

English translation: I don't go with you.

Caló (2): *Las ducais me marelan.* (Bright 1818)

Spanish translation: Las penas me matan.

English translation: The worries kill me.

Basque Para-Romani (*Errumantxela, vascorromaní*)

Errumantxela (1): *Mola pilautzen diat.* (Bakker 1991: 68)

Basque translation: Arnoa edaten diat.

English translation: I drink wine.

Errumantxela (2): *Xaua, goli keauzak, mol buterago akhinen duk.* (Bakker 1991: 68)

Basque translation: Haurra, kantazak, arno gehiago ukanen duk.

English translation: Child, sing, you will have more wine.

2. ROMANI LANGUAGE ON THE IBERIAN PENINSULA