Romani
A Linguistic Introduction

Yaron Matras
University of Manchester

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Yaron Matras 2002

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 10/12 pt. *System* LaTeX 2$_\varepsilon$ [TB]

A catalogue record for this book is available from the British Library

ISBN 0 521 63165 3 hardback

Contents

List of figures *page* viii
List of tables ix
Acknowledgements xi
List of abbreviations xii

1. Introduction 1

2. Romani dialects: a brief overview 5

3. Historical and linguistic origins 14
   3.1 Theories on the origins of the Romani population 14
   3.2 Proto-Romani and Early Romani 18
   3.3 The Romani lexicon 20
      3.3.1 Core and inherited lexicon 20
      3.3.2 Loan components in the inherited lexicon 22
      3.3.3 Semantic domains of the inherited lexicon 25
   3.4 Historical phonology 30
      3.4.1 Changes shared with subcontinental MIA and NIA 30
      3.4.2 Conservative features of Romani 33
      3.4.3 Romani innovations 34
   3.5 Historical morphology 42
      3.5.1 Direct continuation of OIA/MIA productive morphology 42
      3.5.2 Innovations shared with other NIA languages 43
      3.5.3 Romani-specific innovations 45
   3.6 The position of Romani, Domari, and Lomavren 45

4. Descriptive phonology 49
   4.1 Consonants 49
      4.1.1 Stop positions and articulation 49
      4.1.2 Sonorants 50
      4.1.3 Fricatives and semivowels 51
      4.1.4 Affricates 52
      4.1.5 Voicing 53
      4.1.6 Aspiration 54
      4.1.7 Geminates 54
      4.1.8 Consonant clusters 55
      4.1.9 Types of consonant systems 56

6.7 Non-finite forms 159
6.8 Modal expressions 162

7. Syntactic typology 165
7.1 The noun phrase 165
7.2 Constituent order in the verb phrase 167
7.3 Possession and external possession 174
7.4 Complex clauses 176
7.4.1 General features of clause linking 176
7.4.2 Relative clauses 176
7.4.3 Complementation and purpose clauses 179
7.4.4 Adverbial subordination 185
7.5 Negation 189
7.6 The areal position of Romani 190

8. Grammatical borrowing 191
8.1 General considerations 191
8.1.1 Structural borrowing 192
8.1.2 Borrowing hierarchies 193
8.1.3 L2 stratification 194
8.2 Historical layers of grammatical borrowings 196
8.2.1 The pre-European and Greek component 196
8.2.2 Widespread patterns of borrowing following the Early Romani period 199
8.2.3 Rare instances of borrowing 207
8.3 An assessment of grammatical borrowing 209

9. Dialect classification 214
9.1 Methodological considerations 214
9.2 A history of classification schemes 218
9.3 Diagnostic isoglosses 225
9.3.1 Innovations 225
9.3.2 Option selection 229
9.3.3 Simplification 231
9.3.4 Archaisms 234
9.4 Implications of the geographical diffusion model 235

10. Romani sociolinguistics 238
10.1 Aspects of language use 238
10.2 Para-Romani 242
10.3 Romani influence on other languages 249

11. Language planning and codification 251
11.1 Models of language planning 251
11.2 Codification contexts and strategies 254
11.3 Issues of status 257

References 260
Index of dialects 279
Index of names 282
Index of subjects 286

4.2 Vowels 58
4.2.1 Vowel quality 58
4.2.2 Vowel length 59
4.2.3 Other processes affecting vowels 60
4.2.4 Diphthongs 61
4.2.5 Types of vowel systems 61
4.2.6 Stress 62
4.3 Phonological and morpho-phonological processes 64
4.3.1 Historical *ṇḍ* 64
4.3.2 Prothesis and truncation 65
4.3.3 Jotation 67
4.3.4 *s/h* alternation in grammatical paradigms 68
4.3.5 Final *-s* 70

5. Nominal forms and categories 72
5.1 Inherent properties of the noun 72
5.2 Derivation patterns of nouns and adjectives 74
5.2.1 Nominal derivation affixes 74
5.2.2 Nominal compounding and genitive derivations 76
5.2.3 Adjectival derivation 77
5.3 Nominal inflection 78
5.3.1 Case layers 78
5.3.2 Layer I declension classes 80
5.3.3 The independent oblique 85
5.3.4 Forms and functions of Layer II markers 87
5.3.5 Layer III adpositions 91
5.3.6 The stability of synthetic case markers 92
5.4 Adjective inflection 94
5.5 Deictics and related forms 96
5.5.1 Definiteness and indefiniteness 96
5.5.2 Personal pronouns 98
5.5.3 Clitic pronouns 101
5.5.4 Demonstratives 103
5.5.5 The historical development of deictic and anaphoric expressions 106
5.5.6 Other pronouns 112

6. Verb morphology 117
6.1 The basic blueprint for the Romani verb 117
6.2 Verb derivation 119
6.2.1 Word formation and compounding 119
6.2.2 Transitivity and intransitivity 119
6.2.3 The historical development of valency-alteration markers 122
6.3 Loan-verb adaptation 128
6.4 Stem formation and inflection class 135
6.4.1 Present stems 136
6.4.2 Perfective stems 138
6.5 Person concord 143
6.6 Tense–aspect–modality categories 151
6.6.1 Inherited categories 151
6.6.2 Innovations and restructuring 155

Figures

2.1 Location of the principal dialects of Romani *page* 11
2.2 Abbreviations, abstract geographical position, and group affiliation of the principal dialects 12
9.1 Forms for 1. ‘egg’ (**āṇḍa* > **andřo*) and 2. ‘flour’ (**aṭṭa* > *ařo*) 216
9.2 Historical connections among the thirteen Romani dialects according to Miklosich (1872–80, III) 218
9.3 Southeast–northwest division line: *-ipen/-ipe* and *geljas/gelo* 226
9.4 Northwest to southeast cluster reduction 227
9.5 Vlax outwards: affrication in *tikno* > *cikno* ‘small’ 227
9.6 Future-tense marking 228
9.7 Phonological loss of final *-s* 229
9.8 Option selection *s* > *h* and 2SG concord 230
9.9 Selection of 1. stem extension *-in-* and 2. subjunctive *-av-/-ov-* in the copula 231
9.10 Simplification of demonstratives 232
9.11 Simplification of loan-verb adaptation markers: (transitives) 233
9.12 Simplification of perfective concord markers: 1. 2PL 2. 3PL 234
9.13 Retention of definite articles in *-l-* 235

Tables

3.1 Romani innovations shared with Central languages (following Turner 1926) *page* 32
3.2 Retention and shift of historical dental and dental retroflex stops in Romani, Lomavren, and Domari 37
3.3 Principal sound correspondences between OIA and Romani 39
3.4 Some lexical correspondences: Romani and related languages 48
4.1 Proto-Romani consonant phonemes 56
4.2 Early Romani consonant phonemes 56
4.3 Consonant phonemes in Kelderaš and Lovari 57
4.4 Consonant phonemes in Sinti 57
4.5 Consonant phonemes in Gurbet 57
4.6 Consonant phonemes in North Russian (Xaladikta) 58
4.7 Consonant phonemes in Sepeči 58
4.8 Vowels in Arli, Gurbet, Sepeči 62
4.9 Vowels in Kelderaš 62
4.10 Vowels in the the Polska Roma and North Russian (Xaladitka) dialects 62
4.11 Vowels in Welsh Romani 62
4.12 Vowels in Sinti 62
4.13 *s/h* alternation in selected Romani dialects 69
4.14 Treatment of final *-s* in various dialects 70
5.1 Early Romani nominal declension classes, adapted from Elšík (2000a) 83
5.2 Forms of Layer II markers in some dialects (SG/PL) 88
5.3 Early Romani adjective inflection (after Elšík 2000a); athematic singular endings are unstressed. 95
5.4 Forms of the definite article in dialect groups 97
5.5 (Common) Romani personal pronouns 100
5.6 Format layout of demonstrative expressions (with examples) 104

5.7 The four-term demonstrative system in selected dialects (M.SG forms) 104
5.8 Deictic and anaphoric expressions: Proto-Romani Stage 1 106
5.9 Deictic and anaphoric expressions: Proto-Romani Stage 2 107
5.10 Deictic and anaphoric expressions: Proto-Romani Stage 3 109
5.11 Deictic and anaphoric expressions: Early Romani 110
6.1 The basic layout of the Romani verb (with gloss abbreviations) 118
6.2 Reconstructed loan-verb adaptation markers in Early Romani 130
6.3 Loan-verb adaptation markers in Romani dialects 131
6.4 Inflection of Turkish loan verbs in Agia Varvara Vlax: 'to work' < Turkish *çalış-* 134
6.5 Inflection formants of intransitive derivations 137
6.6 Perfective markers: 1SG (jotated) and 3PL (non-jotated) forms 140
6.7 Perfective personal concord markers 145
6.8 The renewal of perfective personal concord in Proto-Romani 148
6.9 The development of perfective personal concord markers (**kard-* > *kerd-* 'done') 149
6.10 Functional arrangement of TAM categories in Early Romani 152
6.11 Early Romani TAM categories and markers 155
6.12 TAM categories and markers in present-day Romani dialects 156
7.1 Linear layout of the noun phrase: principal slots 166
7.2 Adverbial subordinators in Romani 188
8.1 Frequently attested Romani indeclinables by source language 200
8.2 Coordinating conjunctions in some dialects 201
9.1 Classification criteria of the Romani dialects of northeastern Bulgaria according to Gilliat-Smith (1915) 220
9.2 Dialect classification according to Ventcel' and Čerenkov (1976) 221
9.3 Diagnostic features of the Central, Balkan, Vlax branches (based on Boretzky) 224
9.4 Principal differences between Northern Central and Southern Central dialects (based on Boretzky 1999a) 225
10.1 Likely numbers of Romani speakers for some countries 239

1 Introduction

The Rom are known to western culture as nomads and travellers (peripatetics, in anthropological terminology), while to southeastern European society they are familiar as the lowest and most stigmatised social stratum. Stereotypes also surround the image of *Romani*, which is often thought to be synonymous with argot, jargon, or a set of distinct and historically unrelated speech varieties, referred to as 'Gypsy languages'. While there is interface and even some overlap between Romani and argots, just as there is between the Rom and peripatetics, Romani is at its core a language like many others. The agenda of Romani linguistics is consequently similar to that of other fields of investigation in descriptive linguistics: it pursues questions relating to historical reconstruction and structural change, dialect diversification, discourse structure, language maintenance and loss, and more. This book sets out to introduce the structures of Romani and the current agenda of Romani linguistics; parts of it are also an attempt to introduce new ideas into the study of Romani.

Romani is the adjective (feminine singular) derived from *řom*, the historical self-designation of speakers of the language. As a language name, the adjective modifies *čhib* 'language', and so *řomani čhib* means 'language of the *řom*'. It is by far the most widespread term for the language in modern linguistics, and so the most practical cover-term for its various dialects. Speakers can be heard referring to their language as *řomani čhib*, *amari čhib* 'our language', *řomanes* lit. 'in a *rom* way', or by any one of several dozen group-specific names. For lack of any better cover-term for the population of speakers, I shall use the collective form *Rom* – avoiding both the integration into English plural inflection, and the adoption of the Romani plural *Roma* – regardless of individual group affiliation.

Romani-speaking populations are assumed to have settled in Byzantium sometime before the eleventh century (cf. Soulis 1961). References to 'Gypsies' or 'Egyptians' from the eleventh century are believed to relate to them, though we have no definitive evidence that those referred to were indeed Romani speakers. 'Gypsies' then appear in chronicles in other regions, allowing scholars to reconstruct an outwards migration from the Balkans beginning in the fourteenth century, and reaching northern and western Europe in the fifteenth century (Fraser 1992a). Although chronicle references during this period provide

descriptions that match the general image and appearance of the Rom (dark-skinned, organised in family groups, pursuing itinerant trades and especially entertainment), no actual mention of the language is made, nor of their self-ascription. Documentation of the Romani language first appears in the form of wordlists in the early sixteenth century, by which time it is already very close to Romani as we know it today.

The earliest source on Romani is a list of 13 sentences with an English translation, published by Andrew Borde in 1542 under the heading *Egipt speche* (Miklosich 1874–8, IV; Crofton 1907). The State Archives in Groningen contain a manuscript by the magistrate Johan van Ewsum, who died in 1570, with 53 entries of Romani words and phrases accompanied by a Low German translation, under the heading *Clene Gijpta Sprake* (Kluyver 1910). In 1597, Bonaventura Vulcanius, professor in Leiden, printed a list of 53 Romani words with a Latin translation, entitled *De Nubianis erronibus, quos Itali Cingaros appellant, eorumque lingua* (Miklosich 1874–8, IV). The next known sample was collected in 1668 in the Balkans, in western Thrace, by Evliya Çelebi, and published in his well-known travel calendar *Seyāhat-nāme*. It refers to the people called *činganeler* or *qipṭīler*, and contains a brief wordlist and 21 short sentences in their language with a commentary and translation into Ottoman Turkish (Friedman and Dankoff 1991). Job Ludolf's wordlist appeared in Frankfurt in 1691, containing 38 items (Kluge 1901).

The eighteenth century hosted a lively discussion on Romani, and sources are already too numerous to list here. Law enforcement officers in western Europe took a close interest in the speech habits of travellers and minorities. In this context, it was established that Romani and argot (or 'thieves' jargon') were separate linguistic phenomena, and the two were kept apart in compilations such as the Waldheim Glossary of 1727 (reproduced in Kluge 1901: 185–90), the Rotwelsche Grammatik of 1755, and the Sulz List of 1787. In the late 1700s, an international circle of scholars[1] exchanged notes and ideas on Romani, eventually establishing its Indic (Indo-Aryan) origins by comparing it with other languages from around the world. Johann Rüdiger, professor in Halle, was the first to announce the sensational discovery, in April 1777.[2] He then published an article which contained the first grammatical sketch of a Romani dialect, along with systematic structural comparisons of the language with Hindustani (Rüdiger 1782; cf. Matras 1999a). Others followed with similar conclusions (Pallas 1781; Grellmann 1783; Marsden 1785; the latter based on Bryant's list from 1776, see Sampson 1910).

[1] Among them Christian Büttner, Hartwig Bacmeister, Peter Pallas, Johann Biester, and William Marsden; see Pott (1844: 7–16); also Ruch (1986), Matras (1999a).

[2] In his correspondence with his colleague Bacmeister of St Petersburg, though he gives credit to Büttner, who had come to a similar conclusion earlier (Rüdiger 1782: 62; see also Matras 1999a: 95–6; cf. also Ruch 1986: 119–23).

By the time August Pott compiled his comparative grammar and etymological dictionary of Romani (1844–5), he was able to draw on several dozen descriptive sources representing the diversity of European Romani dialects. Pott is usually referred to as the father of modern Romani linguistics, having established the historical and structural coherence of the language and having pointed out the layers of pre-European loan vocabulary, which in turn offered insights into the migration history of the Rom from India to Europe. His book remains the only monograph so far published that is devoted to a comparative and historical discussion of Romani as a whole. Pott's contribution was superseded a generation later, however, by a series of papers by Franz Miklosich (1872–80, 1874–8). This sixteen-part dialectological survey of the language includes a corpus of texts and songs recorded in various parts of Europe, and a comparative and historical grammar and lexicon. By comparing the dialects of Romani, and through the study of selected historical sources, Miklosich was able to reconstruct the migrations of the Rom within Europe, complementing Pott's enterprise.

Two additional landmarks dominate old-generation Romani linguistics. The first is the publication of the *Journal of the Gypsy Lore Society* (1888–; since 2000 under the name *Romani Studies*). However contested some of the social attitudes reflected in its earlier volumes may be, the *Journal* has, since its appearance, served as the principal discussion forum for scientific research on the Romani language as well as a source of data on Romani. The second landmark, closely connected with the *Journal*'s activities, was the appearance in 1926 of John Sampson's monumental grammar and etymological lexicon of the *Dialect of the Gypsies of Wales*, the westernmost variety of Romani, now considered extinct. Alongside these two enterprises, there are numerous other descriptive works from the late nineteenth and early twentieth century that continue to be important and reliable sources of information on the structures of Romani dialects.

Post-war Romani linguistics saw an extension of the research agenda to include issues of language contact and language use, as well as language status and language planning, much of it, during the 1970s and 1980s, embedded into the context of emerging Romani political and cultural activism. A major upsurge of interest in Romani began in the late 1980s and early 1990s, inspired and facilitated by the political transition in central and eastern Europe, where the bulk of the Romani-speaking population lives. The decade from 1990–2000 saw the publication of a large number of monographs, collections, and numerous articles. New fields of interest include grammar, discourse, and typology. During this period, the discipline benefited from funding from national research agencies and governments to promote Romani-related research, from extensive co-operation among specialists working in the field, and from the launch of the International Conferences on Romani Linguistics (first held in Hamburg in 1993).

Recent years have also seen the participation of an increasing number of native speakers of Romani in activities devoted to the study and promotion of their language. Still, the vast majority of linguists specialising in Romani are outsiders to the Romani community. They face the special ethical responsibilities of scholars investigating a society which has not been in a position to produce a scientific tradition of its own. In Europe and urban America, where fieldwork on Romani is typically carried out, such an extreme asymmetrical relationship between the community of investigators and the community that is being investigated is rather exceptional. Ethical responsibility means that one must be cautious of romanticising and of trying to exercise control, but also that one must not be tempted to patronise. Linguistics cannot undo social injustice, nor can it be expected to act primarily in order to promote the self-confidence of Romani communities. There is however a range of services which Romani linguistics can give to the community of speakers, including concrete support of language planning and language education measures. Descriptive linguistics can help replace stereotypical images with information, facts, and evidence.