

## ROMANI

Yaron Matras

Department of Linguistics

University of Manchester

Oxford Road

Manchester M13 9PL

United Kingdom

### Abstract

Romani is an Indo-Aryan language spoken by upwards of 3.5 *Rom* ('Gypsies') in Europe. It shows strong influences of the Balkan languages, especially Greek. Traditionally an oral language, Romani has since the 1990s become a written medium of periodicals and internet sites. In the absence of a unified standard, codification is centred around individual regional dialects.

### Definitions

Romani (referred to by its speakers as *řomani čhib* 'the Romani language' or *řomanes* 'in a Romani way') is the only Indo-Aryan language spoken exclusively in Europe, as well as by emigrant populations in the Americas and Australia. The language is often referred to as 'Gypsy'; it is important however to distinguish between Romani, which is the fully-fledged, everyday family and community language spoken by the people who call themselves *Řom*, and secret or in-group vocabularies employed in various parts of the world, including in Europe, by other populations of peripatetics or so-called service-nomads. There is nevertheless some interface between the two phenomena: In some regions of Europe, especially the western margins (Britain, the Iberian peninsula, Scandinavia), Romani-speaking communities have given up their language in favor of the majority language, but have retained Romani-derived vocabulary as an in-group code. Such codes, for instance *Angloromani* (Britain), *Caló* (Spain), or *Rommani* (Scandinavia) are usually referred to as *Para-Romani* varieties.

In the absence of reliable census figures, the total population of Romani speakers can only be estimated, at anywhere upwards of 3.5 million. The largest concentrations of Romani speakers are in southeastern and central Europe, especially Macedonia, Bulgaria, Romania, and Slovakia. Romani has traditionally been an oral language, and in more

traditional communities there is even opposition to codification attempts or other public use of the language, which is viewed as having protective functions. The overwhelming trend however since the early 1990s has been toward codification of the various dialects at local or regional levels. The language is now used in local media, on numerous internet sites, as a medium of correspondence (especially electronic), and in some countries even as a medium of school instruction.

## History

The earliest attestation of Romani is from 1542, in western Europe. Our understanding of the language's historical development is therefore dependent on reconstruction and comparison with other Indo-Aryan idioms as well as with the contact languages. In phonology Romani shares a number of ancient isoglosses with the Central branch of Indo-Aryan, most notably the realization of Old Indo-Aryan *r̥* as *u* or *i* (Sanskrit *śṛṇ-* Romani *šun-* 'to hear') and of *kṣ-* as *kh* (Sanskrit *akṣi* Romani *j-akh* 'eye'). In contrast however to the other Central languages, Romani preserves a number of dental clusters (Romani *trin* 'three', *phral* 'brother'; cf. Hindi *tīn*, *bhāī*). This had led Turner (1926) to assume a Central origin of Romani, with subsequent migration to the Northwest before the reduction of the relevant clusters took place. A northwestern migration is of course well in line with an ultimate migration out of India and on towards Europe. Further support for Turner's theory comes from the domain of verb morphology, where Romani follows the exact same pattern as Northwestern languages such as Kashmiri or Shina in its renewal of the past-tense conjugation through the adoption of oblique enclitic pronouns as person markers (*kerdo* 'done' + *me* 'me' > *kerdjom* 'I did'). Proto- or pre-European Romani was thus a kind of Indian hybrid: a central Indic dialect that had undergone partial convergence with northern Indic languages. While the retention of dental clusters would suggest a break with the Central languages during the transition period from Old to Middle Indo-Aryan, the overall morphology of Romani indicates that the language participated in some of the significant developments leading toward the emergence of New Indo-Aryan (such as the reduction of the nominal case system to a two-way opposition, nominative vs. oblique, and grammaticalisation of new, postposed case markers). It would appear therefore that Proto-Romani did not leave the Indian subcontinent until late in the second half of the first millennium CE. Romani is among the most conservative New Indo-Aryan languages in retaining a full consonantal present conjugation, as well as consonantal oblique nominal case endings. Typical phonological developments that characterize Romani among the Indo-Aryan languages are the de-voicing of aspirates *bh*, *dh*, *gh* to *ph*, *th*, *kh*, the shift of medial

*d, t* to *l*, of short *a* to *e*, of inflectional *-a* to *-o*, of initial *kh* to *x*, and of the retroflexes *ḍ, ṭ, ḍḍ, ṭṭ, ḍh* etc. to *r* and *ř*.

The subsequent development of the language was strongly influenced by its contact languages. Romani borrowed lexicon and some grammatical vocabulary from Iranian languages and Armenian. The heaviest impact on Early Romani (European Romani, between the 10th and 13th centuries CE) was of Byzantine Greek. Apart from numerous lexical loans, phonemes, and grammatical vocabulary, Romani adopted Greek inflectional morphology in nouns and verbs, which remain productive with loan vocabulary from subsequent European contact languages (see below). Greek also had a strong impact on the syntax of Romani, triggering among other things a shift to VO word order and the emergence of a preposed definite article.

### **The sound system**

Romani dialects generally preserve an aspirated set of voiceless stops *ph, th, kh* as well as *čh*, alongside *p, t, k, č* and *b, d, g, dž*. Nasals are *m* and *n*, fricatives are *f, v, x, h, s, z, š*, and in some dialects also *ž*, and there is an affricate *c* [ts]. All dialects have *l* and *r*, and some also retain *ř*, which is realized as either a uvular [ʀ], a long trill [rr], or in some dialects a retroflex [ɽ,ɽ]. Palatalization of consonants, either distinctive or non-distinctive, is common in the Romani dialects of eastern and southeastern Europe. The vowel system consists of *a, e, i, o, u*, with addition in some dialects of a central vowel *ə* or *ɨ*. Western European dialects of Romani tend to show vowel length distinctions. The phoneme inventory of individual dialects usually accommodates additional phonemes from the respective contact languages in lexical loans. Conservative stress in Romani is on the final inflectional segment of the word, though a number of affixes remain unstressed, among them the vocative ending, agglutinative (Layer II) case endings (see below), and the remoteness tense marker. Dialects in western and central Europe often show a shift of stress to earlier positions in the word.

### **Morphology**

#### *Nominal forms*

Romani nominal morphology is inflectional, with some agglutination. There are two genders, masculine and feminine, and two numbers, singular and plural. Mass nouns often allow omission of overt plural marking. The principal inflectional alternation in the noun is between two ‘basic’ or Layer I cases, nominative and oblique, in the singular and plural. The different patterns of alternation constitute declension classes. Romani declension classes are sensitive to gender, to the phonological shape of the stem, and to etymology

(‘ikeoclitic’ vs. ‘xenoclitic’), with European loans (‘xenoclitic’ nouns) taking Greek-derived case endings. Basic ikeoclitic declension classes are presented in Table 1:

Table 1: Basic ikeoclitic declension classes

	Sg. nominative	Sg. oblique	Pl. nominative	Pl. oblique
Masculines in $-o$	<i>čhav-o</i> ‘boy’	<i>čhav-es-</i>	<i>čhav-e</i>	<i>čhav-en-</i>
Masculines in $-\emptyset$	<i>phral</i> ‘brother’	<i>phral-es-</i>	<i>phral-a</i>	<i>phral-en</i>
Feminines in $-i$	<i>řomn-i</i> ‘woman’	<i>řomn-ja-</i>	<i>řomn-ja</i>	<i>řomn-jen-</i>
Feminines in $-\emptyset$	<i>phen</i> ‘sister’	<i>phen-a-</i>	<i>phen-a</i>	<i>phen-en-</i>

Individual dialects show various patterns of analogies among the different classes. Xenoclitic declension classes typically have Greek-derived inflection endings  $-os$ ,  $-o$ ,  $-is$ , or  $-us$  (masculine) and  $-a$  (feminine), with a variety of plural endings such as  $-i$ ,  $-e$ ,  $-ides$ ,  $-uri$  and more. The oblique stem serves as the base for further (Layer II) agglutinative case formation, with the endings  $-te/-de$  (locative and prepositional),  $-ke/-ge$  (dative),  $-tar/-dar$  (ablative),  $-sa(r)$  (instrumental and comitative), and  $-ker/-ger-$  (genitive). As in other Indo-Aryan languages, the genitive agrees with the head noun (*čhav-es-ker-o phral* ‘the boy’s brother’, *čhav-es-ker-i phen* ‘the boy’s sister’). The oblique without a Layer II extension serves as the case of the direct object (‘accusative’) with animate nouns.

Adjectives usually take vowel endings that agree with the vocalic case-endings of the noun (*mir-o dad* ‘my father’, *mir-i daj* ‘my mother’). Demonstratives usually show a four-term system, encoding both proximity/remoteness (or rather, presence in the situation vs. the discourse context), and general/specific (disambiguation), e.g. *adava*, *akava*, *odova*, *okova*. Interrogatives are cognate with other Indo-Aryan languages (*kon* ‘who’, *kaj* ‘where’), with *so* ‘what’ serving as the base for several derived forms (*savo* ‘which’, *soske* ‘why’, *sode* ‘how many’, etc.). Indefinite markers are often borrowed from the respective contact languages.

### Verbs

Valency is a central feature of Romani verb morphology. It is expressed through direct affixation to the verb root. The productivity, however, of individual valency markers varies among the dialects. Typical valency-increasing markers are  $-av-$ ,  $-ar-$ ,  $-ker-$ , and valency-decreasing markers are  $-jov-$  and  $-áv-$ . They derive verbs from other verb roots, as well as from nouns and adjectives. Borrowed verbs carry loan verb extension or adaptation

markers, based on Greek-derived tense/aspect affixes such as *-iz-*, *-in-*, *-is-*, sometimes in combination with valency affixes (e.g. *-is-ar-*, *-is-ker-*).

The default stem (root with derivation marker) serves as a non-perfective aspect. The plain form of the non-perfective serves as a present/subjunctive. A tense/modality extension *-a* marks the present/indicative, the future, or conditional, depending on the dialect. A perfective aspect (also ‘aorist’ or ‘simple past’) is formed by attaching a perfective extension (derived from the Middle Indo-Aryan participle extension *-t-*) to the root of the verb (e.g. *ker-d-* ‘did’). The choice of perfective extension depends on the numerous perfective classes, which are sensitive to the root phonology as well as to valency and semantics.

There are two person conjugations: The present conjugation (1sg *-av*, 2sg *-es*, 3sg *-el*, 1pl *-as*, 2/3pl *-en*) continues the Middle Indo-Aryan set of present concord markers. There are two inflection classes in the present (non-perfective), distinguishing vocalic and consonantal roots (*xa-s* ‘you eat’, *kam-es* ‘you want’). The perfective conjugation, which follows the perfective extension, derives from late Middle Indo-Aryan enclitic pronouns (1sg *-om*, 2sg *-al/an*, 3sg *-as*, 1pl *-am*, 2pl *-an/-en*, 3pl *-e*).

Both the present and the perfective may be extended by a remoteness marker *-as/-ahi/-ys/-s* that is external to the subject concord marker, indicating the imperfect/habitual/conditional (with the present) or the pluperfect/counterfactual (with the perfective).

## Syntax

Romani stands out among the Indic languages through its Europeanized, specifically Balkanized syntax. Word order is VO, with variation betweenthetic (continuative) VS and categorical (contrastive) SV. Local relations are indicated by prepositions. Adjectives and determiners generally precede the noun, as does the definite article (which agrees with the noun in gender, number, and case). Relative clauses are postposed, and often introduced by a universal relativizer *kaj* < ‘where’. Clauses are generally finite. Adverbial clauses are introduced by conjunctions, usually derived from interrogatives. Romani distinguishes between factual and non-factual complex clauses. Modal, manipulation, and purpose clauses are introduced by a non-factual conjunction *te*, as are conditional clauses. Epistemic complements are introduced by *kaj*, which is often replaced by a borrowing from the respective contact language.

### **Dialect diversity**

Dialect differentiation in Romani appears to have emerged largely in situ, following the dispersal of groups from the Balkans into western and northern Europe, from around the 14th century onwards, and their settlement in their present locations, during the 16th-17th centuries. There are two major diffusion centres of innovations: in the southeast, especially the northern Balkans, and in western-central Europe, especially Germany. Typical of the western-northern dialects are prothesis of *j-*, simplification of *ndř* to *r*, loss of adjectival past-tense in intransitives (*gelo, geli* > *geljas* 'he/she went'), and retention of *-n* in the abstract nominalizer *-ipen/-iben*. In the central regions, *s* in grammatical paradigms is often replaced by *h*. Individual regions show distinct developments in morphological paradigms, especially demonstratives, 2/3pl perfective concord markers, and loan verb markers. Especially these latter isoglosses justify the current classification into the following dialect groups: Balkan (with a sub-group 'Black Sea Coast'), Vlax (Transylvania and adjoining regions), Central, Northeast (Baltic-Northrussian), and Northwest (German-Scandinavian) (see Maps 1 and 2).

### **Bibliography**

- Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Bakker, P. & Matras, Y. (2003) *Bibliography of modern Romani linguistics*. Amsterdam: John Benjamins.
- Elšík, V. & Matras, Y. (eds.) (2000). *Grammatical relations in Romani: The noun phrase*. Amsterdam: John Benjamins.
- Turner, R. (1926) 'The position of Romani in Indo-Aryan'. *Journal of the Gypsy Lore Society* (3)5, 145-189.