

## 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 *řom* 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 *qiptīler*, 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<sup>1</sup> 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.<sup>2</sup> 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).

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

<sup>2</sup> 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).

4 Introduction

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.

## 2 Romani dialects: a brief overview

---

The present chapter provides a brief overview of the principal dialects of Romani that have been described in the linguistic literature, focusing in particular on the dialects that are cited in the following chapters. It does not pretend to offer a complete survey of dialect names or locations; for additional references to dialects of Romani see the list of dialects in Bakker and Matras (1997: xxiv–xxvi) and the dialect index in Elšík and Matras (2000: 229–32).

Speakers usually refer to their language as *romani čhib*, *romanes* ‘Romani’ or as *amari čhib* ‘our language’, or else derive the term from the individual group designation, using either a genitive compound – *lovarengi čhib* ‘the language of the Lovara’ – or an adverbial derivation – *sintitikes* ‘the Sinti way (of speaking)’. In the descriptive literature, dialects are often referred to using either the group name in the plural – ‘the *Xaladitka* dialect’ –, or reinterpreting the name as a singular – *Bugurdži* lit. ‘drill-maker’, *Sinto* lit. ‘a Sinto’, *Arli* lit. ‘settled person’. Terms for a single dialect may differ when two distinct groups speak dialects that are close enough to be considered one and the same by linguists. On the other hand, terms may overlap when two communities speaking distinct dialects share a name based on their religious affiliation, trade, or region of origin. In addition, internal designations used by groups often differ from external designations applied to them by other Romani-speaking populations.

There are several types of group names in Romani. A number of groups simply refer to themselves as *rom*, or use other specific ethnic designations such as *romaničel*, *kale*, *manuš*, *sinte* (cf. Wolf 1960a; see also chapter 3). This is the conservative pattern, and the one more widespread in western and northern Europe. In the Balkans and central-eastern Europe, group designations may be based on traditional trades, the actual terms being borrowed mainly from Turkish, Romanian, or Hungarian: *bugurdži* ‘drill-makers’ (Turkish *burgucu*), *sepeči* ‘basket-weavers’ (Turkish *sepetçi*), *kelderara/kelderaša* ‘kettle-maker’ (Romanian *căldărar*), *čurari* ‘sieve-maker’ (Romanian *ciurar*), *lovari* ‘horse-dealer’ (Hungarian *lo-v-* with a Romanian-derived agentive suffix), *ursari* ‘bear-leader’ (Romanian), and many more.

The distinction between itinerant Rom and settled Rom is highlighted in some group names (cf. Paspatis 1870). A widespread term in the southern Balkans

is *erli/arli* from Turkish *yerli* ‘settled’, used to denote mainly Muslim settled populations. Some groups associate themselves with the nation among which they have settled, often using a general term for non-Roma as an attribute: *gačkene sinte* ‘German (< *gadžikane* ‘non-Romani’) Sinti, *xoraxane rom* ‘Turkish/Muslim Rom’ (< *xoraxaj/koraxaj* ‘foreigner’). Many designations are more specific, denoting country of settlement – *polska roma* ‘Polish Rom’ –, the region of settlement – *bergitka roma* ‘mountain Rom’ (of the southern Polish highlands) –, the place of origin – *mačvaja* ‘from the district of Mačva in Serbia (a group based in the United States, Russia, and Sweden) – or, as an external designation, the (often mistakenly) assumed origin – *romungri* ‘Hungarian Rom’ (Polish and Russian Rom as referred to by Lovara).

Since the following chapters refer to the structures of varieties of Romani as described by linguists, it seems preferable to repeat the terminology used by the individual authors. As a reference grid I shall be using in part the recent division into dialect groups, as outlined and employed in Bakker and Matras (1997), Bakker (1999), Elšík (2000b), Matras (2000a) and Boretzky (2001) (see also chapter 9).

We begin with the historical centre of Romani population diffusion, in the **Balkans**. The Romani dialects of the southern Balkans (Turkey, Greece, Bulgaria, Macedonia, Albania, Kosovo) are generally referred to as the ‘Balkan’ branch, which in turn is divided into two groups. The more conservative, southern group includes the Rumelian sedentary dialect described by Paspatis (1870); the dialects of the Sepečides or basket-weavers of northern Greece and Turkey (Cech and Heinschink 1999); the dialects known as Arli or Arlije, which are spoken in Greece, Albania, Macedonia, and Kosovo (Boretzky 1996a), one of the major dialects of the region in terms of numbers and geographical distribution of speakers; the Erli dialect of Sofia, documented by Gilliat-Smith (1944, 1945; cf. Calvet 1982, Minkov 1997, Boretzky 1998a); the dialect of the Crimean Rom (Toropov 1994), which nowadays is spoken mainly in Kuban’ and Georgia; the Ursari dialect spoken in Romania (Constantinescu 1878); and the dialects of Prilep (Macedonia), Prizren (Kosovo), and Serres (northern Greece), which are Arli-type but considered by Boretzky (1999b) as separate varieties. Recent work in Greece has documented additional dialects, some of them with very conservative features: the dialect of the *romacel* musicians, called *romacilikanes*, of the Ipeiros district (A. Theodosiou p.c.), an additional and distinct dialect of Serres (I. Sechidou p.c.), and the dialect of Pyrgos in the Peloponnese (N. Christodoulou p.c.). The conservative Balkan group also includes a number of closely related dialects spoken in northern Iran, which are clearly European dialects of Romani whose speakers migrated eastwards: the dialect of the Zargari in Azerbaijan (Windfuhr 1970), and the dialect called Romano in northeastern Iran (Djonedj 1996).

A second group within the Balkan dialects emerged in northeastern Bulgaria. They are referred to in the following as the **Drindari–Kalajdži–Bugurdži group**; Boretzky (2000b) has referred to them as Southern Balkan II. The group includes the dialect of the Drindari (also known as Čalgidžis or Kitadžis) of Kotel and Varna in northeastern Bulgaria (Gilliat-Smith 1914; also Kenrick 1967), the dialect of the Kalajdži tanners of Tatar Pazardžik, Bulgaria (Gilliat-Smith 1935), as well as what appear to be immigrant dialects in Macedonia and Kosovo, such as that described by Uhlik (1965) for Skopje, and the Bugurdži (or Rabadži) dialect described by Boretzky (1993a).

Both Balkan sub-groups are characterised by a continuous Greek influence that appears to have lasted longer than the Greek influence on dialects that left the southern Balkans, as well as by a strong Turkish influence. Many speakers of the Balkan dialects are Muslims, and many retain active knowledge of Turkish. Speakers of Arli varieties in particular, from Macedonia and Kosovo, are also found in western Europe, especially in Germany and Austria, where they settled as labour migrants or asylum seekers between the 1960s and 1990s, as well as in the United States.

Probably the most ‘prominent’ group of Romani dialects – in terms of numbers of speakers, geographical distribution, and the extent of documentation – is the **Vlax** branch. It is believed that Vlax emerged in Romanian-speaking territory. The dialects share extensive Romanian influence on vocabulary, phonology, and loan morphology, as well as a series of internal innovations. There were many migration waves of Vlax speakers from the Romanian principalities, some of them at least connected with the abolition of serfdom in Romania, which lasted until the second half of the nineteenth century. The branch is split into two groups.

The **Southern Vlax** dialects are documented mostly for migrant communities that have settled outside Romanian-speaking territory. The Southern Vlax dialects of Valachia/Muntenia (Constantinescu 1878) and of northeastern Bulgaria (Gilliat-Smith 1915) are closest to their original locations. Farther south, there are two divisions.

In the southeast, we find the Southern Vlax varieties of Greece. Some were spoken by Christian nomadic groups during the nineteenth century (cf. Paspatis 1870). Others are spoken by Christian immigrants from Turkey who were resettled in the 1920s. These are known as Kalpazea, Filipidžía, and Xandurja. Large communities are reported in Dendropotamos near Thessaloniki (Tong 1983) and in Athens; the only thoroughly described variety is spoken in the district of Agia Varvara in Athens (Iglá 1996).

In the southwest, we find dialects generally referred to in the literature as the ‘Gurbet-type’, based on the group name *gurbet* employed by some. Other names include *džambazi* and *das* ‘Slavs’. Unlike the speakers of Balkan Romani

dialects among whom they live, the Gurbet-type varieties are spoken mainly by Christians. Descriptions and documentations exist for Serbia and Bosnia (Ackerley 1941, Uhlik 1941 and elsewhere), Albania (Mann 1933, 1935), and Kosovo (Boretzky 1986). There are however also Muslim groups of speakers, such as the migrant group in Italy, which calls itself *xoraxane* ('Muslims') (Franzese 1986).

The **Northern Vlax** sub-branch includes two dialects on which we have fairly extensive documentation. The first is the dialect of the Kelderaš (or Kalderaš), which, alongside (Balkan) Ursari, is probably the most widely spoken Romani dialect in Romania. It has numerous sub-divisions, with names usually reflecting the very intact clan structure that exists among the group. An extensive text documentation and comments on grammar of the Bukovina dialects is included in Miklosich (1872–80, iv–v). Detailed grammatical descriptions of Kelderaš are based exclusively on migrant dialects: Gjerdman and Ljungberg (1963) for a variety spoken in Sweden, Boretzky (1994) for a dialect of Serbia, Hancock (1995a) for an American contact variety of Mačvaja (Serbian Northern Vlax) and Russian Kelderaš, and Matras (1994) for a contact variety of Lovari and Kelderaš originally from Transylvania, spoken in Poland, Germany, and Sweden. There are large communities of Russian Kelderaš speakers in Argentina and Brazil.

The second is the dialect of the Lovari, formed in Transylvania in contact with Hungarian. Lovari is now the main variety of Romani spoken in Hungary (e.g. Mészáros 1968). Lovari groups had already migrated into Austria and Germany in the nineteenth century (Ackerley 1932). Other communities have settled in Slovakia, Poland, Yugoslavia (Vojvodina), and Scandinavia. Descriptive outlines of Lovari include Pobožniak (1964) for southern Poland, and Cech and Heinschink (1998) for Austria. Recent collections of Lovari narratives are Gjerde (1994) for Norway, and Cech, Fennesz-Juhász and Heinschink (1998) for Austria. There are other Northern Vlax dialects, such as Čurari, which are not very well described. A recent CD-collection of songs and narratives in various Vlax dialects of Hungary and Romania is available in Bari (1999; cf. also Bari 1990). A further dialect of Hungary, Cerhari (Mészáros 1976), represents a transitional variety, sharing a number of diagnostic features with both the (Northern) Vlax and the Central dialects. Also affiliated with the Vlax branch, but with some independent developments, are the dialects of southeastern Ukraine (Barannikov 1934).

The **Central** branch of Romani dialects is also divided into two groups. The **Northern Central** dialects include the now extinct Bohemian Romani (Puchmayer 1821), West Slovak Romani (von Sowa 1887), and East Slovak Romani (Hübschmannová et al. 1991). The latter is now the dominant variety in the Czech Republic, due to the massive immigration of eastern Slovak Roma to Bohemia in the late 1940s to early 1950s, and is the variety most

widely used in text production in this country. Northern Central dialects are also spoken in southern Poland (Rozwadowski 1936, Kopernicki 1930), Moravia, and Transcarpathian Ukraine. The Northern Central dialects retain a layer of Hungarian influence.

The **Southern Central** dialects are sometimes referred to as the *-ahi* dialects due to their characteristic imperfect/pluperfect suffix. They are further sub-divided into two groups. The first, eastern, group is collectively known as Romungro ('Hungarian Rom'). In Hungary itself, Romungro is only spoken by a very small number of speakers, following a large-scale shift to Hungarian. Documentation includes Görög (1985). Other Romungro dialects are spoken in Slovakia (Elšík et al. 1999). The second, western, group is known as the Vend group, and includes dialects of western Hungary (Vekerdi 1984), the Prekmurje variety of northern Slovenia (Štrukelj 1980), as well as the Roman dialect spoken by the Rom in the Burgenland district of Austria (Halwachs 1998). All Southern Central dialects show considerable Hungarian influence. The Gurvari dialect of Hungary (Vekerdi 1971a) is a transitional variety which has absorbed many Vlax influences.

Several diverse dialect groups and individual varieties are sometimes referred to collectively as a '**Northern**' branch, although they are spoken not only in the north of Europe but also in the west and extreme south. 'Northern' will be used in the following chapters primarily in citation. Instead, the groups and isolated dialects will be referred to individually. In the centre of the so-called 'Northern' branch we find the closely related Sinti-Manuš varieties. They all share strong German influence and a number of innovations, and it seems that the group emerged in German-speaking territory, with sub-groups migrating to other regions. The first grammatical outline of a Romani dialect, by Rüdiger (1782), was devoted to a Sinti variety. There is extensive documentation of short texts and narratives in various German Sinti varieties, almost all from the pre-war period. Grammatical descriptions of German Sinti varieties include Liebich (1863), Finck (1903), and most recently Holzinger (1993, 1995). Closely related to German Sinti is the dialect of the Manuš of France (Jean 1970, Valet 1991). German Sinti varieties are also spoken in the Netherlands, Austria, as well as in Hungary (Vekerdi 1983), Bohemia, Slovakia, Russia, and Yugoslavia. There is in addition a southern branch of Sinti in northern Italy: the rather conservative Piedmontese Sinti (Franzese 1985), Lombard and Venetian Sinti (Soravia 1977), and the varieties of the Sinti Estrexarja or Austrian Sinti of South Tirol (Tauber 1999). It appears that *Manuš* and *Kale* are the older names used by the groups, whereas *Sinti* first appears in the eighteenth century (cf. Matras 1999a:108–12).

Related to Sinti is the Finnish dialect of Romani (Bourgeois 1911, Thessleff 1912, Valtonen 1972, van der Voort 1991, Koivisto 1994), which has only a very small number of speakers, perhaps just a few thousand. From historical records, and from the Swedish element in the dialect, it is clear that the Finnish

Rom or Kaale migrated via Sweden. The series of features that are shared with Sinti allows us to speak of a **Northwestern** group, with a historical centre in German-speaking territory. In the other Scandinavian countries, traces of Romani (apart from Vlax-speaking immigrant communities) remain only in the special vocabularies used by peripatetic populations (Etzler 1944, Iversen 1944, Johansson 1977). A dialect once spoken in northern Estonia by the Rom of Laiuse, or Lajenge Roma, now appears to be extinct (Ariste 1964), following the persecution and annihilation of most speakers under the Nazi occupation. While sharing some features with the neighbouring Baltic dialects, it has strong connections to Finnish Romani and the Northwestern group, including Swedish influences, which suggest that the dialect was once part of the Finnish sub-group.

A fairly coherent dialect branch is the Polish–Baltic–North Russian or **Northeastern** group. Best documented is the North Russian or Xaladitka dialect (Sergievskij 1931, Wentzel 1980). Closely related to this dialect is the dialect of central Poland, spoken by a group who refer to themselves as the Polska Roma (Matras 1999b). Latvian Romani, also known as the Čuxny dialect (a Russian term for Estonians) or as Lotfiko/Loftiko, is spoken by a small population in Lithuania and Latvia as well as in Estonia (Mānušs 1997; Kochanowski 1946). Little documentation exists on a further Baltic dialect, once spoken in eastern Latvia and Lithuania (Ariste 1964).

**British** Romani, an independent branch, is now considered extinct. The most thorough and extensive description is Sampson's (1926) monumental grammar of Welsh Romani or the Kååle dialect, which was still spoken by a number of families until the second half of the twentieth century (cf. Tipler 1957). English Romani appears to have become extinct towards the end of the nineteenth century, and survives only in the form of a special lexicon. Both forms of English Romani, termed the 'old' and the 'new' dialect, are described by Smart and Crofton (1875). It is possible that the oldest documentation of a Romani dialect by Borde in 1542 (see Miklosich 1874–8, iv; Crofton 1907) is based on British Romani.

**Iberian** Romani is also extinct, and survives only as a special lexicon in Spanish-based Caló (< *kalo* 'black'; Bakker 1995, Leigh 1998) and Basque-based Errumantxela (< *romaničel*; Ackerley 1929, Bakker 1991). Sources from the nineteenth century however allow us to reconstruct fragments of the variety of Romani that was spoken in Catalonia (Ackerley 1914).

Finally, there are two rather isolated groups of dialects. The first are the dialects of southern Italy–Abruzzian and Calabrian Romani (Soravia 1977) and Molisean Romani (Ascoli 1865). They are strongly influenced by Italian, and appear to be early offshoots of the Balkan dialects. The second is the Croatian dialect, for which there is no documentation from Croatia itself. Speakers of the dialect in Slovenia refer to themselves as Dolenjski Roma (i.e. from the lower province of central Slovenia), while a sub-group in Italy call themselves