

Matras, Yaron. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press. 2002. xiv+291 pp. £50.00.

*Romani Studies* (Series 5) Vol. 13, No. 2, 2003. 163-173.

Although not all Gypsies speak Romani, all native speakers of Romani as their first community language are Gypsies (in a restricted, ethnic sense of the term). For some groups of Gypsies who do not speak Romani, the evidence is clear that Romani was the primary community language at one time and that the community has subsequently undergone language shift. The term *Gypsy*, however, is sometimes also used to refer to peripatetic social groups who may or may not speak or have ever spoken Romani. In the case of some West European groups to whom the term *Gypsy* is applied, it is certain that Romani is not and was not used as a community language. Similarly, the so-called ‘Gypsy languages’ of India, described, for example, by G. A. Grierson in volume XI of the *Linguistic Survey of India* (1922, Calcutta, Office of the superintendent of government printing), are not forms of Romani but Indic languages of India that have been separated from Romani for at least as long as all the other languages of India (more than a millennium) and whose speakers occupy socio-economic niches similar to those of some Gypsies in Europe. In some recent West European historical and anthropological literature, a confusion has arisen between the use of *Gypsy* as an ethnic label to refer to Romani-speaking or formerly Romani-speaking groups (who may or may not be peripatetic) and the second usage, which is more social than ethnic and does not have a firmly definable linguistic basis. Members of some groups of *Gypsies* in this second sense have secret lexicons that are not independent languages. Some historians and anthropologists have become so confused by these two usages of the term *Gypsy* as to assert that Romani is merely lexicon, not a language, and that there was no migration of speakers of Early Romani to Europe from India but rather that peripatetic peoples passed Indic lexical items from one group to another along trade routes. Yaron Matras’ *Romani: A Linguistic Introduction* (henceforth RLI) is a milestone in the past two centuries of linguistic scholarship on Romani that has demonstrated conclusively that Romani is indeed a language, that it is Indic, and that the only reasonable hypothesis for its presence in Europe is that a group or some groups of speakers migrated thither from India. But M’s book is not really concerned with refuting the claims of a few misguided western scholars unacquainted with the field of linguistics, although the material presented in it does so. Rather, this volume gives both a state-of-the-art survey of current knowledge relating to the Romani language and introduces some of the author’s original ideas concerning various aspects of Romani linguistics. RLI is not only the first such work to be published in English but is also the best of its kind in the field. It is essential reading for anyone concerned with Roms or the Romani language, but it also has much to contribute to general linguistics and can be used with profit by scholars in other areas of the humanities and the social sciences.

The book is divided into eleven chapters: 1 Introduction (1-4), 2 Romani dialects: a brief overview (5-13), 3 Historical and linguistic origins (14-48), 4 Descriptive phonology (49-71), 5 Nominal forms and categories, 6 Verb morphology (117-164), 7 Syntactic typology (165-190), 8 Grammatical borrowing (191-213), 9 Dialect classification (214-37), 10 Romani sociolinguistics (238-50), and 11 Language planing and codification (251-59), followed by a bibliography of over 450 references (260-78) and indices of dialects (279-81), names (282-85), and subjects (286-91). Chapters Four through Seven constitute the core of the synchronic description of the structure of Romani combined with relevant diachronic explanations. These chapters are aimed specifically at linguists and require at least a basic knowledge of linguistics. The theoretical orientation is functional-structural, and the material and analyses are accessible to any linguist, clearly and coherently presented, and contain a number of new and convincing explanations. The first three chapters and the last four are concerned with those aspects of Romani linguistics that are of interest to both linguists and scholars in other fields. They can be assigned independently of Chapters 4-7 in any course for which Roms in particular or Gypsies in either sense constitute the focus of study or a relevant subject.

Chapter One introduces Romani and gives a brief overview of the history of its study, mentioning all the earliest accounts up to the eighteenth century and then surveying the most important figures and events up to the present. The author concludes this chapter by addressing the ethical responsibility of researchers to the community that is the subject of research and assessing in a realistic fashion the contributions that Romani linguistics can make to the Romani-speaking community.

Chapter Two gives a very brief overview of the dialectal classification of Romani currently accepted by most linguists working on Romani. The four living branches are Balkan, Vlax, Central, and Northern, each of which is divided into two subgroups — southern and northern for the first three and

eastern and western for the last. M also mentions British/Welsh and Iberian Romani, each considered as forming independent branches and surviving only as special lexicons embedded in territorial languages. M also treats the Abruzzian and Hravati/Dolenjski dialects as taxonomically distinct, although it appears that they are basically southern Balkan dialects that separated early from the main group and subsequently underwent significant changes under the influence of other centers of innovation and territorial languages. I shall return to this issue in my discussion of Chapter Nine, to which M refers the reader in his mention of these dialects. Chapter Two concludes with a geographic and a schematic map of the main dialects that serve as the bases of discussion in RLI and a brief mention of Para-Romani, the Romani-based lexicons that function as special variants of territorial languages and that are discussed in Chapter Ten. Owing to the dialectal diversity of Romani and the fact that it is spoken all over Europe (and, as a result of subsequent migration, on most other continents as well), particularly by some populations that are or were peripatetic, Romani is described as a non-territorial language, although specific dialects are associated with geographic locations. Any linguistic study of Romani must take into account its dialectal diversity, since there is no dominant standard that can serve as a point of reference. Chapter Two thus serves as the basic frame for the rest of the discussion. Since RLI is indeed an introduction and not an exhaustive compendium, and, moreover, since the documentation of Romani dialects and their nomenclature is the subject of on-going research (most notably the Romani Morpho-Syntactic Database Project at the University of Manchester directed by Yaron Matras and the Romani Lexical Database Project directed by Dieter Halwachs of the University of Graz, Peter Bakker of the University of Aarhus, and Yaron Matras of the University of Manchester), RLI does not attempt to provide a complete account of all Romani dialects but rather an inclusive overview. Thus, for example, Figure 2.1 *Location of the principal dialects of Romani* leaves Albania completely blank not because Romani is not spoken there, nor because we lack documentation for some of the local dialects, but because the principal dialectal types needed to form the basis of discussion in RLI are represented by other well-documented dialects spoken in adjacent countries.

Chapter Three surveys Romani historical linguistics and its contribution to the reconstruction of Romani history. M's sober and careful account makes it clear that while historical linguistics establishes the origins of Romani in India beyond the shadow of a doubt, and moreover gives us some indications of the time and route of migration, it cannot tell us the reason for that migration, nor can it establish the social position of the speakers whose language became Romani. As M points out, the ethnonym *Rom* (řrom/) is certainly cognate with the attested caste name *Dom*, which refers to various service-providing populations in India and is also cognate with ethnonyms of Indic-speaking or formerly Indic-speaking commercial nomads (some of whom use other names) who live all over central and western Asia and adjacent parts of Africa from Tajikistan and the Caucasus to Egypt and Sudan. Having discussed all the theories of Romani's social origin, M concludes that what we know for certain is that mobile populations of Indic origin specializing in certain trades and retaining distinct ethnic and linguistic identities were present in the Near East before the Roms arrived in Byzantium "sometime around the eleventh century or earlier." (18). Worth citing here is evidence from Tzitzilis (2001), who argues on the basis of the phonology and morphology of Middle Greek loan-words that Romani was being spoken in Byzantium by the tenth century. Note particularly that Tzitzilis (2001:327) observes that the Greek source of Romani *kurkó* is not the usually cited *kyriakí* but a different expression for 'Sunday' *kyrikón (ímar)* (pace p. 205). M defines *Proto-Romani* as the Indic dialect that became a distinct language in India and *Early Romani* as the language that was spoken in contact with Greek in Byzantium prior to the dispersal of Romani-speaking populations throughout Europe beginning in the thirteenth or fourteenth centuries. These are distinct from *Common Romani*, which is the term he uses for those Early Romani elements shared by attested Romani dialects, as opposed to variations that may have been present in Early Romani or Proto-Romani. In his discussion of the Early Romani inherited lexicon, M identifies about 1,000 lexical roots, of which 200-250 are from Greek, around 70 from Iranian, about 40 from Armenian, leaving about 650-700 roots of Indic origin plus a few items from Georgian, Ossetian, and, perhaps, Burushaski. M discusses some of the most significant examples as well as their semantic domains (kinship terms, body parts, etc.) and the question of what this lexicon can tell us about Romani culture. After a short but careful survey of conflicting hypotheses, M concludes that words relating to areas of contact with surrounding populations are more likely candidates for loan-replacements while those dealing with more intimate spheres show greater stability and resistance to replacement. He cites Boretzky's (1992) conclusion that Romani's relatively small common inherited vocabulary flags resistance to assimilation. M gives a fairly detailed account of Romani historical phonology, a brief survey of the most important features of historical morphology, and a discussion of the relationship of Romani to Domari and Lomavren. He supports

Turner's (1926) arguments for the origin of all three languages in Central India. M also notes that Turner failed to take into account the Romani preservation of a distinct reflex for retroflex \**d*, and, citing the reflexes of Old Indic *bhrātr̥*- 'brother' (Romani *phral*, Lomavren *phal*, Domari *bar*) along with other phonological, morphological, and lexical evidence, allows for the possibility that the three languages became distinct as a result of their emergence as non-territorial languages in Central India, with subsequent shared innovations resulting from areal contact as the groups followed similar paths of migration through Northwestern India while perhaps separating completely upon reaching Iranian-speaking territory.

The descriptive linguistic core of RLI, chapters Four through Seven, discuss phonology, the nominal system, the verbal system, and syntactic typology. The opening treatment of palatals in Chapter Four is a bit confusing, since /č/ is described as behaving "differently from *other* stops" (emphasis mine), but later in the paragraph M points out the need to distinguish palatal stops, palatalized consonants, and affricates. It would have been better to formulate this description in terms of four primary positions for *obstruents* (labial, dental, palatal, velar), three of which always have stops. Otherwise the chapter gives a thorough survey of Romani consonantism, vocalism, and prosody as well as the main morphophonemic alternations: Prothesis/truncation, jotation, and the alternation of /s/ with /h/, /j/, or zero. The treatment of this last alternation gives a much needed synthesis of a complex phenomenon and is based in large part on M's own research. For this alternation, which actually involves two phenomena — an older one in intervocalic position and a younger one in final position (the former restricted to certain grammatical morphemes, the latter either grammatical or phonological depending on the dialect) — M establishes a hierarchy of occurrence and argues that the older phenomenon can be traced back to variation in Proto-Romani and Early Romani beginning with the 2 singular present tense marker (-*esa*) and spreading by analogy first to the 1 plural present, then the instrumental singular, the copula, (3 singular present before other forms), and interrogatives, in various dialects which tend to cluster in Central Europe. Loss of final /s/ in Greek-derived nouns, the 3 singular past marker (-*as*), the reflexive marker (*pes*), masculine accusative (independent oblique) singular (-*es*), adverbial -*es*, the short present (2 singular, 1 plural), and inherited lexical items represents different processes in different dialects and also follows a hierarchy of occurrence. The most conservative are those at the periphery.

Chapter Five treats the inflection and derivation of substantives, adjectives, and deictics (pronouns and articles). In discussing case markers, M employs Masica's (1991) terminology: Layers I, II, and III. In Masica (1991:231-32, 234, 236), Layer I markers are "bits of material inherited from M[iddle]I[ndo-]A[ryan]/O[ld]I[ndo-]A[ryan]," Layer II markers are "attached to the base indirectly, through the mediation of a Layer I element and/or invariant for all nouns and the same for all numbers." Layer III markers are mediated by Layer II, and, when they are not, are still distinguishable by their lack of morphophonemic variation, ability to be polysyllabic, and, usually, a "fairly transparent connection with an independent word." Masica needs this system owing to the fact that the Indic languages of India are all postpositional and therefore the interface between the three layers requires delicate distinctions. M's use of Masica's terminology is justified, for the most part, on historical grounds, since most of the Romani substantival inflectional system derives from the innovations of the Middle Indic period that led to the formation of the first two layers. In Romani, however, the material corresponding to Masica's Layer III consists of prepositions which are always identifiable as independent words and frequently do not trigger the occurrence of Layer II markers (except in pronouns or in some dialects in contact with languages having complex case inflection). Moreover, the Romani vocative markers, as M notes (p. 80), connect directly to the base but are of late formation and do not belong to Layer I (nor to Layer II). M's treatment of the independently occurring oblique stem (frequently labeled "accusative") as "encoding the **non-agentive referent** that is **high on the topicality scale**" (p. 86) captures the fact that this form is not limited to animate or definite direct objects but also occurs in other functions. M's account of the development of Romani's distinctively complex set of deictics is original and convincing.

Chapter Six gives a thorough account of Romani inflectional and derivational verb morphology. M analyzes the four basic paradigmatic sets in terms of two oppositions: Remote/non-remote and perfective/non-perfective. The remoteness marker is -*as* (*sine* in the Arli of Kosovo and Macedonia) and the perfective stem uses the participial base and person markers of pronominal origin (see below). The four paradigms thus defined are the present (-R -P), imperfect (+R -P), preterite or aorist (-R +P), and pluperfect (+R +P). The pluperfect can also function as a counterfactual. M has a third non-remote category, *intentional*, which he labels as marking the subjunctive, e.g. long *kerava* 'do' 1 sg present vs short *kerav* 1 sg subjunctive. While this may have been the case in Early Romani, such a description is problematic in the modern dialects that do not use the long form as the future. Although it is common for

the short present to occur after modal markers such as the future marker *ka* and the modal subordinator *te*, it is also the case in dialects such as Skopje Arli that the long form occurs after *te* when introducing the protasis of a conditional period as well as after *bi* marking the apodosis, e.g. *Te čingarea man, me bi avava* ‘If they were to invite me I would come’. Short forms can also occur in simple statements, e.g. *E Rifatos pendžarav, e čhaja da pendžarav* ‘I know Rifat and I know his daughter’. Consider also the following radio announcement, which has long and short indicatives: *O Ajnuri thaj o Džemo tar-i Švedska bahtaren e pranden e Ramijeske thaj e Mirsadake a e Safeteske thaj e Sadijake bahtarena o bijav* ‘Ajnur and Džemo from Sweden congratulate Rami and Mirsada on their marriage, and they congratulate Safet and Sadija on their wedding.’ Thus the characterization in Table 6.12 (p. 156) needs further explication. M’s explanation for the origin of the simple preterite in an ergative-like construction going back to a past participle plus a possessive or relative linking particle *\*jo* plus a pronoun is original, ingenious, and convincing. It gives the first truly satisfactory account for both jotation and person-marking in the preterite.

Chapter 7 illustrates the most important syntactic features of Romani using 96 example sentences, all taken from unelicited data sets. M’s account details the syntax of noun phrases, constituent order in verb phrases, possession, complex clauses, and negation. The discussion of Romani syntax brings together much varied material, and this is perhaps the most complex and difficult area of Romani to survey. Thus, for example, Macedonian *deka* is also used as a factive complementizer in Arli, and could in fact have been the source for Greek *oti*, which also occurs in Macedonian (cf. p. 179), Balkan Turkish *eger* ‘if’ is used alongside South Slavic *ako* in Macedonia (cf. p. 187), etc. M makes a useful distinction between stable, Common Romani Balkanisms and those Balkan contact features that are specific to the dialects of Southeastern Europe. In the conclusion, he identifies three Early Romani innovations as typical of Romani syntax: 1) relative clauses with resumptive pronouns and interrogatives as relativizers; 2) a non-factive/factive distinction in subordinators (native *te* for the former in almost all dialects, native *kaj* [‘where’] or non-native forms for the latter); 3) predominant VO word order with OV for focus, VS for connective-narrative, and SV for contrastive thematic orders.

Chapter Eight, discusses Romani in the context of the field of contact linguistics. The origins of Romani are quite distinct from what is normally understood by the term *contact language*, i.e., a language coming into existence as the result of contact between two or more speech communities. Rather, as M observes, Romani is a language ‘permanently in contact’ (p. 191) with other languages. There are no monolingual speakers aside from very young children, and until the twentieth century there were no organized efforts at language preservation. While there are other speech communities in the world where bi- or multilingualism is the universal, Romani is, perhaps, unique in its history of non-territoriality combined with what must be assumed to have been centuries of universal multilingualism. M discusses Romani evidence in the context of various theoretical attempts to predict the relative borrowability of grammatical elements. Using Romani data he establishes an implicational hierarchy of the likelihood of borrowing in conjunctions (from least to most frequently borrowed and/or replaced): ‘and’ < ‘or’ < ‘but’. Other areas of borrowing are discourse markers, subordinating conjunctions of reason, cause, factivity, and time, various expressions of modality, adjectival gradation, bound indefinite markers, focal quantifiers, and bound morphemes used in nominal derivation. In nominal inflection, M mentions the borrowing of the nominative plural markers from Romanian (*-urj-/uri*), Greek (*-ides*), and Greek/Slavic *-i*. He also cites the use of *-e* in the Balkans as being of South Slavic origin, but it is unclear why this could not be an extension of the plural marker for native masculine nouns in *-o*, albeit possibly with South Slavic reinforcement. We can also mention here the influence of Slavic inflectional patterns on adjectival agreement, which M notes on p. 96. Borrowed prepositions generally express abstract rather than concrete spatial relations. M also notes a variety of phonological influences from contact languages (mainly stress and vowels) as well as syntactic convergences. Among those elements which are not borrowed and thus represent significant constraints are the following: Demonstratives, bound tense markers (unless attached to native forms), productive definite articles (except on borrowed prepositions), bound case markers, personal and possessive pronouns, locative ‘in’, interrogative ‘what’, numerals below ‘five’ as well as ‘ten’ and ‘twenty’, and, except in Dolenjski, the copula, and replacements for the modal subordinator (non-factive complementizer) *te*. Romani displays morphological compartmentalization in distinguishing inherited from borrowed elements. As M notes, this is especially characteristic of Romani contact phenomena, allowing it to preserve an inherited core while permitting massive borrowing. (Cf. Friedman 2001, which reaches a similar conclusions on the split between areas of Balkan Romani open to contact influence and those resistant to it.) At the same time, M notes that the

semantic-pragmatic nature of borrowed categories suggests a cognitive motivation: Categories concerned with the internal structure of meaning (deixis, case, tense, aspect, non-factive *te*) are more resistant to borrowing than categories expressing the external structure of meaning (discourse markers, various particles and conjunctions [especially contrastive], modality, aksionsart, word order, and the factive complementizer). M also discusses hierarchies of borrowing within the realm of categories likely to be borrowed.

Chapter Nine gives a detailed account of Romani dialect classification. M gives a thorough history of the most important classification schemes and a lengthy discussion of diagnostic isoglosses, including mappings of 14 features. He gives the details for the currently accepted four-branch/two subgroups apiece model of living dialects, but emphasizes the usefulness of viewing dialects in terms of option selection and geographical diffusion to account of the complications of all the relevant data, i.e. he favors the wave model over the tree (branching) model. M identifies three main centers of diffusion of innovations: Vlach and Sinti, which create a southeast/northwest division, and Balkan. He finds the geographical diffusion model more useful in understanding the position of peripheral dialects such as Hravati/Dolenjski. The position of such ‘problem’ dialects can be compared with that of, e.g., the southernmost dialects of Serbian, which are transitional to Macedonian and Bulgarian but whose oldest distinctive features are shared with the Serbian core. In the South Slavic case, as in the Romani one, dialects are formed by competing processes of divergence and convergence. Older innovations (or their absence) indicate the position of a dialect vis-à-vis older divisions and centers of change, while later developments can emanate from other centers. Nonetheless, in both cases it is possible to construct a taxonomy assigning a systemic position, if only for heuristic purposes. Be that as it may, M’s original contributions and coherent synthesis constitute the best and most up-to-date general account of Romani dialect classification currently available.

Chapter Ten on Romani sociolinguistics discusses the problem of estimating the actual number of Romani speakers, issues of bilingualism, code-switching and code-mixing, Para-Romani, and the Romani element in other languages. M makes the important point that the tendency for Romani words to be recruited into slang, especially in the expression of subjects that are pejorized or taboo in mainstream society, is an expression of the high prestige of Romani for anti-establishment, marginalized communities. As slang itself is a challenge to established norms, the Romani element in slang indicates its significance for challenges to the social order. On the other hand, the final chapter of RLI treats Romani as a participant in the established social order, namely through codification and language planning. M covers the main issues of this question and shows that the Romani standardization process is decentralized and pluralist. It is worth noting that the situation in Macedonia, where efforts at creating a Romani standard go back to the early 1970’s, current trends in language planning belie traditional linear or even circular models of the relevant processes and require more complex analyses. M concludes with a discussion of the status of Romani and notes the rise in its recognition in academic research contexts and the importance of Romani web sites in increasing Romani’s presence in the public domain. To this we can add that the achievement of M’s book itself is a significant event in the advancement of the status of Romani.

RLI demonstrates that M has both a thorough mastery of Romani and a firm grasp of a broad range of theoretical concerns. The book will be of tremendous use to a broad range of scholars — not only linguists but also social scientists and other students in various humanistic disciplines. There is nothing like it available in any language, and it fills a pressing need. RLI’s theoretical frameworks are sound, the explanations are well argued, and the coverage is thorough. M discusses every major topic of concern to Romani and of relevance to Romani’s relation to a broad range of theoretical linguistic issues. This book is not merely a collection of already established knowledge but an original contribution to the advancement of both Romani scholarship and a variety of theoretical concerns within linguistics.

#### References

- Boretzky, Norbert. 1992. Zum Erbwortschatz des Romani. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 45: 227-251.
- Friedman, Victor A. 2001. Romani Multilingualism in its Balkan Context. *Sprachtypologie und Universalienforschung*, 54.146-159.
- Masica, Colin. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University.
- Turner, Ralph L. 1926. The position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society*, 5(3rd series).145-189.
- Tzitzilis, Christos. 2001. Mittelgriechische Lehnwörter im Romanes. “Was ich noch sagen wollte...” A *Multilingual Festschrift for Norbert Boretzky on the Occasion of His 65th Birthday (Sprachtypologie und*

*Universalienforschung, Supplements, Studia typologica 2*), ed. by Birgit Igla and Thomas Stolz, 328-340. Berlin: Akademie Verlag.