Mapping the Romani dialects of Romania

Yaron Matras (University of Manchester)
2013

http://romani.humanities.manchester.ac.uk/virtuallibrary

EUROPEAN COMMISSION
European Research Area

SEVENTH FRAMEWORK PROGRAMME

Funded under Socio-economic Sciences & Humanities

Mapping the Romani dialects of Romania

YARON MATRAS

This is the first attempt to map specifically the Romani dialects of Romania – the country with the largest Romani-speaking population. The data under consideration derives from fieldwork carried out over the past decade as part of the dialectological documentation of the Manchester Romani Project, based on the Romani Dialectological Questionnaire, first drafted in 2001. Most of the data has been made accessible online via the Romani Morpho-Syntax Database. The analysis focuses on key features that have been cited in the past as distinctive of internal (rather than contact-induced) historical developments in the individual dialects of Romani. The results show that Romania constitutes, on the one hand, a zone of ongoing structural innovations that is prone to random diversity for some forms. On the other hand, in respect of some developments, distinct geographical diffusion zones can be identified. These tend to mirror topographical and historical political boundaries. These patterns provide further evidence that Romani dialects are best viewed as a geographical continuum that mirrors historical contacts between Romani populations in situ at the time in which structural change took place. Thus, they call into question the notion of 'genetic' divisions among dialect groups and their speaker populations.

Keywords: Romani language, Romania, dialect, linguistics, Vlax, Central

1. Introduction

In Western perception there is a strong affinity between *Roma* and *Romania*. This derives partly from the fact that Romania is home to the largest Romani-speaking population, estimated at anywhere between 500,000 and 2,500,000. In part it derives from the popular impression, first created through a migration wave of Romanian Roma westwards during the second half of the nineteenth century and again since 1990, that Roma have their origin in Romania. The accidental similarity of names contributes further to this association. Yet

Yaron Matras is the Editor of Romani Studies and Professor of Linguistics at the School of Arts, Languages and Cultures of the University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom. Email: yaron.matras@manchester.ac.uk

despite the considerable interest in Romanian Roma outside of Romania and despite their conspicuous presence within the country, there have been surprisingly few attempts to document the Romani dialects of Romania. In fact, not a single modern (post-1940s) work exists that is devoted to the descriptive study of Romani as spoken within the territory of Romania. This stands in sharp contrast to a proliferation of pedagogical publications devoted to the teaching and learning of Romani that have appeared in this country since 1990, and the publication in Romania of several Romani dictionaries and of creative writing in Romani. Linguistic–philological essays tend to be confined to the documentation of oral tales (e.g. Gaster 1931) and to the discussion of Romani borrowings in Romanian, especially in Romanian argot (Graur 1934, Juilland 1952, Drimba 1992, Leschber 1995, Bochmann 1999). A re-print of Constantinescu's (1878) collection of texts, which appeared in 2000, is quite possibly the only accessible descriptive monograph devoted to Romani in Romania.[1]

The absence of documentation goes hand in hand with the absence of analyses of the Romani varieties of Romania. Perhaps the most intensive discussion of Romanian Romani varieties is included in Boretzky's (2003) historical and descriptive outline of the so-called 'Vlax' (Vlach) dialect group (on classification schemes of Romani dialects see below). The nomenclature approach in this and other works by Boretzky is predicated on an agreement-in-principle on a pre-determined set of features that define the group or 'branch'. Examples of structural features that are used to define the Vlax group are the historical change in the vowel representing the 1SG past-tense marker *-om > -em*, the presence of the expression *khanči* 'nothing', the addition of a prothetic *v-* to the set of third-person pronouns, as in *ov > vov* 'he', the use of either *ni* or *či* as negation markers with finite lexical verbs, and more. These in turn are used as a basis for the postulation of an even more comprehensive inventory of shared forms, whose territorial spread is then considered only in a secondary step of the analysis. In the meantime, other neighbouring and co-territorial varieties that do not share the pre-determined set of distinctive structural features are left outside the scope of the investigation. I shall return to a critique of this method below.

The paucity of original data on Romanian varieties of Romani that has so far been available is evidenced by Elšík and Matras's (2006) study, which covers over seventy varieties of Romani, none of them from Romania (cf. pp. 421–2). In the past few years, however, this gap has been covered through fieldwork undertaken as part of the Romani Dialectological Database Project based

---

1. Boretzky and Igla's (2004) dialect atlas of Romani takes into consideration data from altogether five sources from Romania, three of which are classified as 'Vlax', one as 'Ursari' and one as 'Spoitori' (though in some of the maps the dialect described as 'Kalajdži of Vidin (Bulgaria)' is depicted within Romanian territory, e.g. 2004, Vol II: 317–18, 320).

at the University of Manchester. The online database[2] now offers the most comprehensive documentation of Romani dialects in general, including data obtained from well over one hundred different sources (i.e. speakers in various locations). The following is an attempt to assess the degree of variation found among these samples without prejudice to the attribution of structures or clusters of structures to any particular 'dialect branch'.

My underlying assumption is that structural variation in related languages and dialects is the outcome of structural change in language. Change is introduced by a sector within the speech community and is gradually propagated throughout the community, just like any other cultural fashion. In this way linguistic changes can spread from one family to another, or from one age group to another, within a language community in a particular location (Labov 1994). Change can also spread further, to other individuals in other locations, as a result of contacts and imitation of fashionable speech forms among individuals in different locations. Social, political and geographical barriers to contact between individuals and communities may constitute obstacles to the spread of innovative language use. As a result, the linguistic map can be read as a map of socio-historical contacts among population units (families, communities, groups in certain locations, etc.). Tighter linguistic resemblance will represent a history of close contacts and mutual or unilateral accommodation, while linguistic differences will represent either absence of contact or social barriers that prevent accommodation.

2. A brief history of dialect classification schemes in Romani

The concept of 'dialects of Romani' refers to the related speech forms of dispersed populations, who usually refer to their language as *romanes* or related terms (*romnes*, *romaneh*, *romacilikanes*, and so on). The shared inventory of structures that connects these speech varieties has been well defined for over a century (see Miklosich 1872–1880, Sampson 1926) and has a proven history of having developed out of related speech varieties that were spoken in medieval India (see already Pott 1844; cf. Turner 1926, Matras 2002, Elšík and Matras 2006). While one cannot of course exclude the possibility that the speaker communities continued to absorb members from outside the population of shared descent after their settlement in Europe, it is at the same time obvious (pace sceptics such as Okely 1983, Lucassen 1996, and Willems 1997) that the core of this population, at the very least, descend from migrants that arrived from India, and who spoke an Indo-Aryan language. It is thus clear that there is only partial overlap between the community of speakers of Romanes (Romani)

2. http://romani.humanities.manchester.ac.uk/rms

and the entire population referred to in popular perception as 'Gypsies and Travellers' and in the ethnographic literature as 'peripatetics' or 'commercial nomads'. That social boundaries are upheld between those that are and those that are not speakers of Romani, is taken for granted. The question pursued in the present context is whether language can shed light on the boundaries and connections among individual population groups within the dispersed speech community of Romani, that is, what language can tell us about historical contacts among Romani-speaking populations.

Inevitably, our discussion touches on previous attempts to sub-divide the dialects of Romani. Such attempts are a common academic exercise in the study of dialects of any language. They arise from an effort to understand the role that structural changes have had on shaping the linguistic landscape. Quite often, the spread of changes can be correlated with social barriers of various kinds to obtain a picture of the historical diffusion and dispersion of what had once been a uniform population, or in turn of the history of merger or convergence of populations that had once been further apart. In this respect, the study of dialect divisions and dialect classification is often linked to the study of historical political landscapes (as in the case of the dialects of England), of a way of life (e.g. urban, rural and nomadic, as in the case of dialects of Arabic), and of population movements.

In Romani dialectology, there are several reasons to emphasise migrations in connection with dialect diversity. Firstly, the notion of migration accompanies our understanding of the very origin of Romani and the history of its presence in Europe. Next, it has been shown that the history of successive migrations has left its mark on the language in the form of layers of contact influences, representing accommodation processes to other languages at different times and in different places. The successive nature of pre-European foreign influences such as Iranian, Armenian, and Greek is disputable, and it is highly possible that all three lexical layers may have been part of a synchronous multilingual input in medieval central and eastern Anatolia (see Matras 2002).[3] But the later development paths of individual European dialects of Romani certainly do illustrate a diverse history of contacts. The earliest conceptualisation of the emergence of dialect splits in Romani (Miklosich 1872–1880) therefore depicted a gradual migration wave through Europe, from which individual groups broke away and became subjected to local influences as well as to internal processes of change.

Capturing movement across both time and geographical space, Miklosich's idea of sub-branching from a main wave of migrants was in many ways akin to the predominant tree-diagram representing sub-branching over time (and

---

3. Armenian and Greek were both spoken in this area until the early twentieth century, while (Iranian) Kurdish continues to be one of the dominant languages of the area, and knowledge of Persian, once the principal lingua franca, is widespread to this day.

possibly also in space) that had been used in the nineteenth century, and is still in use to this day, to depict language-genealogical relations. Miklosich's model resulted in the postulation of dialect groupings that more or less overlapped with the majority population of a particular region, for example 'Hungarian Romani' and 'German Romani'. Implicit in this terminology is the notion that the contact language plays a key role in shaping individual dialects, but also that cultural contacts with a certain population will consolidate the Romani group and erect boundaries that help constrain internal change, boundaries that are in turn instrumental in blocking influences from neighbouring dialect groups. In other words, once a Romani population (which we might envisage as a clan or group of related clans) settles in a region, its customs and practices, both linguistic and non-linguistic, absorb influences from the surrounding settled population. By adopting local customs and local linguistic influences such as vocabulary and possibly structural features as well, the particular Romani population develops its own cultural and linguistic distinctiveness, setting it apart from other Romani groups. The particular group-identity that emerges as a result may serve to limit contacts with other Romani groups and to inhibit influences from other Romani populations. In its strict formulation, however, Miklosich's model remains focused on the historical 'sub-branching' of population groups as the key to understanding the accumulation of group-particular structural changes.

Even more explicitly focused on the notion of 'branch' as detached from territoriality is Gilliat-Smith's (1915) discussion of the Romani dialects of north-eastern Bulgaria. Gilliat-Smith identified two linguistic groups among the Roma of the region, whose settlement patterns overlapped and who were only partly differentiated through attributes such as religion (Christian Orthodox or Muslim), occupation pattern (nomadic or settled), and major contact language (Bulgarian or Turkish). He concluded that one of the groups, labelled Vlax, had immigrated from a Romanian-speaking territory and had spread in an area that had already been inhabited by local (non-Vlax) Romani speaking communities. In the decades that followed, Romani communities speaking a variety of dialects that were to some degree related to Gilliat-Smith's Vlax (and in some cases also referred to themselves, or were referred to by others, as Vlaxo/Laxo etc.) were recognised throughout Europe. The notion of Vlax migrations – in all likelihood in several waves, originating from different regions and at different historical times – took on a major position in the overall conceptualisation of Romani-speaking populations. The contrast between Vlax and non-Vlax has since been applied consistently to label local Romani dialects across Europe. The dialect classification model promoted largely in educational and activist publications by Marcel Courthiade (1998) offers a kind of blend between Miklosich's approach and that of Gilliat-Smith:

Present-day dialects are depicted in the form of two distinct historical layers of population migrations. The first spans the greater part of Europe, having spread historically via the Balkans to central Europe and from there onwards in different directions. The second has its point of departure in the Romanian (Wallachian and Moldavian) principalities, spreading in most directions through later (post-1800) migrations.

The difficulty with this kind of division is that it neglects the considerable diversity of the 'older', or non-Vlax layer of dialects, and promotes the very particular innovations that constitute the Vlax group to the major division line among Romani dialects. In order to do justice to a more balanced model, labels have been suggested for various sub-groups of dialects, which, impressionistically, were regarded as closely related in the sense that they shared structural features and were perhaps also more easily mutually comprehensible. Models of this kind were put forward by Vencel and Čerenkov (1976) and Kaufman (1979), paving the way to what has later been termed a 'consensus' classification, recognising the Vlax, Balkan, Central, and Northern dialect groups. First adopted in passing by Boretzky and Igla (1993), this classification was somewhat modified and described in more detail by Bakker and Matras (1997) and subsequently by Elšík and Matras (2006).

While the latter two works merely treat the framework as a reference grid,[4] Boretzky (1999a, 1999b, 2003, 2007) has in various works argued for the 'genetic' coherence of the individual groups as dialect 'branches' of Romani. Boretzky's method consists primarily of delimiting the group under discussion in the first instance through a pre-selection of sources, and then taking an inventory of the features found within each corpus of sources. Ironically, this enumeration of features is usually carried out using maps representing the territorial spread of structures, but limiting the examination of their spread to the pre-selected group. Boretzky and Igla's (2004) atlas of Romani dialects offers a synthesis of this information, and demonstrates quite clearly that many isoglosses in Romani are far-reaching and encompass more than the individual groups and sub-groups identified on the 'consensus grid'. In spite of the picture that the maps deliver, the scenario of historical dialect differentiation that is adopted in the atlas, and later more elaborately in Boretzky (2007), is this: The major dialect groups formed within Romani during the migration process into Europe on the basis of pre-existing differences among the speech varieties of different clans. These differences, acquired before and during the migration process, became further pronounced as the various population groups settled in different areas, leading to the formation of 'dialect groups'.

4. But see Bakker (1999) for arguments in favour of a 'genetic' grouping referred to as the 'Northern' branch, and a critique of the 'Northern' branch in Matras (1999).

The assumption is therefore that dialect divisions come about through migrations of separate groups, and that they are in that sense largely legacies retained from a remote location, or 'inherited' rather than acquired in situ. This assumption seems to be motivated by a tradition that emphasises the connection between language and migration in Romani philology – the realisation that Romani arrived in Europe with a migrant population, Miklosich's idea of groups breaking away from a major migrant wave, and Gilliat-Smith's realisation that the spread of Vlax dialects is a result of a secondary (internal European) migration. It is furthermore strengthened by a pre-conception of 'Gypsies' as a nomadic population, detached from territory. However, based on a comparison of present-day Romani dialect forms it is quite straightforward to reconstruct a more or less uniform 'Early Romani' which will have been spoken in Anatolia during the Byzantine period (see Matras 2002, Elšík and Matras 2006). The dialect divisions as we see them today can therefore be assumed to have emerged following the immigration of Romani-speaking populations into Europe and their settlement in the various regions, not before.

The assumption that dialect boundaries reflect a division into groups that existed prior to migration and settlement in Europe is not borne out by the overall pattern of main isoglosses within Romani, either. In Matras (2002, 2005, 2010: Ch. 2) I pointed out that major isoglosses take on different shapes and paths in the geographical landscape. In many instances, we find a split between centre and geographical periphery, which cannot possibly be accounted for by the model that attributes innovations to the random migration patterns of pre-existing, coherent population groups. In other cases, isoglosses tend to cluster around what I have termed the 'Great Divide' – a zone that represents the historical border area between the Habsburg and Ottoman Empires. Many other isoglosses encompass smaller regions, dividing Europe from west to east and from northeast to centre. Only the combination of a series of salient isoglosses, primarily those pertaining to the morphology of paradigms (such as loan verb adaptation markers, person endings, and demonstratives), yields a picture that resembles that of the consensus grid. In other words, the consensus grid is inspired by a selective impression of differences among the dialects at the level of morphology, in the first instance.

The resulting hypothesis is that a more or less uniform Early Romani spread into Europe in historically attested waves of Romani immigration during the late fourteenth and early fifteenth centuries. The groups then settled in the various European regions, gradually abandoning Greek as a second language and replacing it with the respective majority languages of their new regions of settlement, and accommodating to some of the local customs while establishing regional economic networks. The dispersal thus led to a partial segregation of groups. It is hypothesised that the major changes that separate the dialects

took place during this period. Some changes, such as the shift to initial word stress in central and western European dialects of Romani, will have been a direct result of the new linguistic contacts. Other changes will have been internal innovations. Changing social norms will have accelerated the propagation of innovations within individual communities. Written attestation of Romani becomes quite dense by the early eighteenth century, and by this time the dialects are known to have acquired a shape that is very similar to the one that is recognisable today. It follows that the major dialect differences emerged in situ during the sixteenth and seventeenth centuries. The timing fits in well with the bundle of isoglosses that constitutes the Great Divide, which historically mirrors what was a border and a conflict zone during this very period: It constituted an obstacle to contacts among Romani groups across the political boundary of the two large Empires and prevented northern innovations from diffusing south, and southern innovations from spreading north.

3. Data collection and data processing methods

The data assessed in this article were collected between 2003 and 2010 by fieldwork assistants and research collaborators[5] with access to Romani speakers in Romania, using the dialectological questionnaire developed by the Manchester Romani Project in 2001.[6] The questionnaire is designed specifically to capture morpho-syntactic and other structural variation (lexical, lexical–phonological) among the dialects of Romani. It is typologically informed and includes a comprehensive and systematic elicitation survey of the various domains of morpho-syntax, from personal pronouns and indefinites through to various adverbial, complement and relative clauses and on to an overview of local and temporal relations and more. The individual semantic categories are accommodated into phrases and sentences, which are read to speakers in the respective majority language of the country. Speakers are asked to translate them orally into their variety of Romani. The list also includes individual words and full conjugations of verbs representing the various Romani verb classes. Elicitation usually takes between four and five hours and is accompanied by recordings of free conversation, mostly descriptions of festivities and customs. The material is recorded digitally (earlier analogue recordings have been digitised) and added to an archive of recordings, accompanied by relevant metadata on the place of the recording, the speaker's life history, occupation, knowledge of other languages, group self-ascription, religious practices, and more.

6. See http://romani.humanities.manchester.ac.uk/rms/browse/phrases/phraselist

Recordings of speakers' translations of the questionnaire phrases are tran-scribed onto a spreadsheet template in which each phrase is pre-tagged for the structural and semantic categories that it is anticipated to contain: from examples of key phonemes, through to inflection endings, temporal and local expressions, types of complex constructions, conjunctions, and so on. The transcriptions are then imported into a database, in which the tags are cross-referenced to table cells representing the various structures. During the editing process, the editor of a given table retrieves relevant example phrases for each individual table cell by opening up an example window for a given sample (i.e. data obtained from an individual speaker source). The data are entered into the table based on an assessment of the example sentence. Once the editing process is complete, the user is able to view information organised in tables via a web browser. By clicking on a table cell, the user can retrieve the example sentences. The audio recording is segmented and cross-referenced with the phrases, so that the user may also listen to the original speaker pronounc-ing the phrase. Finally, through a separate procedure, extracts of connected speech are also assessed and phrases are tagged for relevant categories. The user browsing through tables on structural information can in this way also obtain access to relevant phrase examples obtained from connected speech. The user can listen to the phrase and be re-directed to its original position within the full conversation transcript. The Romani Morpho-Syntax Database[7] thus forms a state-of-the-art resource that delivers a comprehensive structural documentation of varieties of Romani from more than one hundred differ-ent locations, accompanied by example phrases from both natural and elicited speech in transcription and original audio. The database is equipped with a search facility, and users can create and download tables for comparison and plot search results on an interactive map.

The data for the present survey were obtained from the Manchester Romani Dialect Database and accompanying archive. The maps were re-drawn to represent the specific picture in Romania.[8] Individual samples each represent the data obtained from one single speaker. These are correlated via a serial number to the location and the self-appellation given by the speakers for their respective groups (see Table 1). The governing principle behind this system of reference is that dialect names are not static and therefore they do not provide stable, unambiguous reference labels. Firstly, group endonyms and exonyms may change over time or even vary according to the conversational setting or the value attributed to a group at a given moment as a result of status or alliances (cf. Marushiakova and Popov 2013). Location names are not a reli-able indicator of the structural profile of a dialect, either, since dialects can be

TABLE 1. Dialect samples, locations, and speakers' self-designations

Sample number	Location	Group self-designation
RO-001	Cluj	Romungro
RO-004	Maglavit	Ursarja
RO-006	Piteşti	Spoitorja
RO-008	Piteşti	Kəldərarja
RO-009	Valureni	Gabor
RO-010	Crăciuneşti	Gabor
RO-011	Timişoara	Gabor
RO-012	Târgu Mureş	Šušuwaje
RO-013	Huedin	Piculesči Roma
RO-015	Huedin	Kurturare
RO-016	Călăraşi	Spoitorja
RO-019	Corneşti	Gabor
RO-020	Sadova	Ursarja
RO-021	Bolentin Vale	Spoitorja
RO-025	Deaj	Kərəmidarja
RO-026	Budiu Mic	Gabor
RO-031	Crăciuneşti	Gabor
RO-036	Sangeorgiu de Mureş	Žuržuvaje
RO-048	Spinus-Oradea	Roma Ungrika
RO-049	Dej	Rom Mătase
RO-050	Tinca	Kərəmidari
RO-051	Velt	Čurarja
RO-052	Vereşti-Hancea, Suceava	Kelderari
RO-053	Albeşti, Mureş	Rom Mătase
RO-054	Caransebeş, Timiş	Ungrika Rom
RO-055	Lugaşu de Jos, Bihor	Kaštale
RO-056	Nasaud	Teglari
RO-058	Senereus	Vlaši Roma
RO-059	Bahnea	Romungro
RO-060	Albeşti, Mureş	Rom
RO-061	Diosig, Bihor	Lovari
RO-062	Ineu de Cris	Roma Lingurari
RO-063	Săcuieni	Ungrika Rom
RO-064	Ţăndărei	Kangljari
RO-065	Timişoara	Kalderaša
RO-071	Uileacu de Cris	Čurari
RO-072	Glodeni	Romungro
RO-073	Lugaşu de Jos, Bihor	Kaštale
RO-074	Sadova	Kərəmidarja
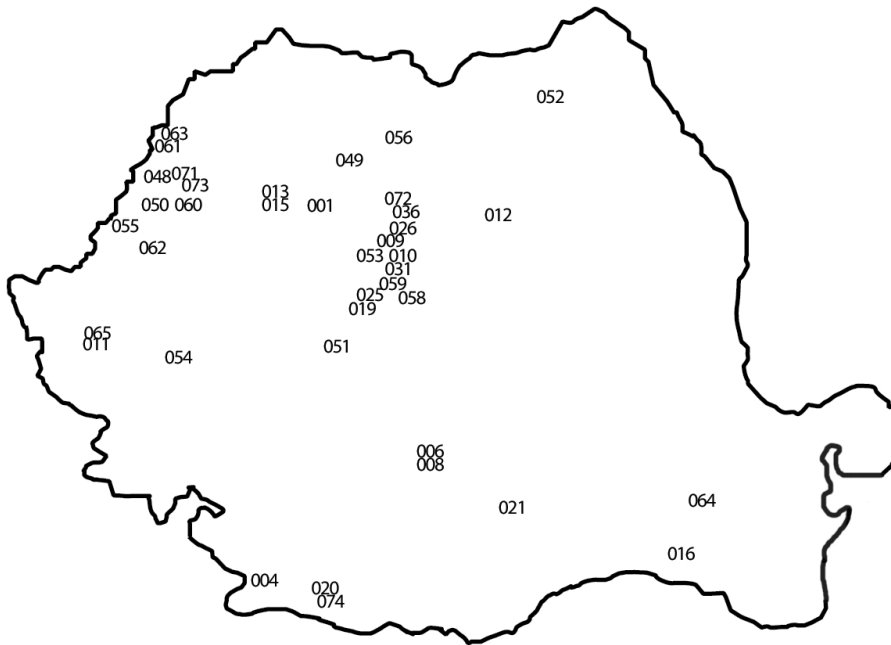
MAP 1. Locations of sample numbers

shared across several neighbouring or even distant locations, while an individual location may host two or more groups that speak different dialects. There is therefore a need for an independent, objective reference system, and this is achieved by assigning each sample obtained from a single speaker a numerical code, which is preceded by the respective country code (in this case 'RO' for 'Romania'). The speaker's place of birth and residence and the group's self-designation that is provided by the speaker in response to the question "What do you call your Romani group?" are recorded and filed as attributes of the sample reference. The extended inventory of attributes includes the additional ethnographic data about customs, celebrations, occupation profiles and relations with other groups (see Beluschi Fabeni, in this issue).

Map 1 shows the geographical distribution of the samples. As can be seen with reference to the self-designations listed in Table 1, only some groups tend to show geographical clustering: while the Romungro, Ungrika Rom or Roma Ungrika are settled primarily in Transylvania, the Ursarja are in the south of the country, and the Spoitorja are in the southeast, the Gabor have their centre in Transylvania but branch out to Banat, while the Kalderaša/Kelderari/Kəldərarja and the Kərəmidarja appear to be dispersed randomly across the country. The question of the internal coherence of the dialects of groups who carry the same or related self-appellations is discussed in section 6 (see also Urech and van den Heuvel 2011).

4. Features of interest in the Romani dialects of Romania

The Romani linguistic landscape in Romania is interesting due to its diversity. In terms of the 'consensus' classification grid, Romania is home to dialects belonging to at least four distinct so-called 'dialect branches': the Balkan branch, represented by the Ursari dialects in the south and along the Black Sea coast; the so-called Balkan-*zis*-Branch (South Balkan II in Boretzky's terminology; 1999b),[9] represented by the Spoitori dialects of the south; the Vlax branch, represented by the continuum of Kelderaš, Kurturari, and other related varieties; and the Central branch, represented by some of the Transylvanian and Maramureş varieties, often referred to as Romungro. Still unclarified in relation to the 'consensus' grid is the status of the Romani dialects of the Republic of Moldova, which have only recently been documented (Manchester Romani Project, RMS Database). In relation to the macro-space geographical analysis of isogloss distribution (Matras 2002), Romania is situated at the southern edge of Great Divide, while at the same time it is at or near the centre of diffusion for some innovations.

The question that confronts us in this, as in other contexts of Romani dialectology, and for which the Romanian area can serve as an interesting indicator, is whether a diverse dialect zone is one in which pre-formed dialects can be assumed to have come into contact with one another as a result of migration and dislocation from their original, formative areas, as in the case described by Gilliat-Smith (1915) for northeastern Bulgaria, or whether the mixed zone is in fact a transition zone that accommodates more or less fuzzy and possibly even transient isoglosses reflecting the reach of various innovations from different centres of diffusion. Of interest in this connection are competing forms, that is, those forms that lack uniformity across the region, and which at the same time represent features that are shared, respectively, with different dialects and dialect groups outside the geographical zone under scrutiny. Romanian influences are ruled out of the consideration since their appearance within the zone can easily be attributed to contact with Romanian as a contiguous language, a factor that is consistently at play for all dialects concerned. Equally beyond consideration are macro-geographical developments that fail to reach the area of southeastern Europe as a whole, such as the prothesis of *j-* in words like *jov* 'he', *javer* 'other' and so on, or the dropping of initial *a-* in *mal* 'friend', *čh-* 'stay', both features that are found exclusively north of the Great Divide.

In order to tackle head-on the question of dialect group and isogloss formation in Romani, it makes sense for us to select those features for examination that appear in the Romanian Romani linguistic landscape and at the same time

9. The division between South Balkan I and II seems to imply that these are viewed as subdivisions of a single 'branch'.

represent variants that are often attributed as diagnostic features to the vari-
ous so-called dialect 'branches'. We are concerned especially with competing
variants that appear to have a consistent distribution to the north or west of
Romania, and to the south, respectively, but which occur in variation or along-
side one another within Romanian territory itself, as well as with innovations
that appear to have their origin within Romania itself (but are not a direct
result of Romanian influence). These can be said to constitute the features of
the isogloss bundle that crosses Romania – the southeastern edge of the Great
Divide.

When comparing dialect forms, it is vital to distinguish archaisms from
innovations. The retention of archaisms is due to the absence of change. While
some coherent regions may appear to be resistant to change, the retention of
conservative features in different locations does not necessarily point to a shared
development, that is, it is not necessarily the outcome of contacts among the
speaker populations but merely an indication of a shared legacy. By contrast,
innovations involve spread rather than retention, and so shared innovations
within a coherent region are likely to represent networks of historical social
contacts among speaker populations. This means that a dialect-geographical
analysis must go hand in hand with an historical reconstruction of structural
developments. For Romani, reconstruction poses a special challenge due to
the absence of written documentation on earlier forms of the language. Instead
it relies heavily on a comparison of dialects, on an understanding of general
mechanisms of change, and on a comparison with related languages of India
(for an in-depth discussion of the method see Matras 2002).

A conservative feature that is found exclusively in some, though not all, of
the Romani dialects of southeastern Europe (including out-migrant dialects
from the region) is the continuation of the historical consonant cluster *ṇḍ
(Proto-Romani) or *ndř* (Early Romani) in words like *manřo* 'bread' as a cluster
with the variants *nř*, *nr*, *rn*, *ndr*, *ngr*, and more, or else as a distinctive uvular
or retroflex trill (/ʀ/ or /ṛ/). Another phonological archaism preserved in the
region is the retention, in the form of the definite article (usually oblique and
plural forms *le-*, *la*), of the consonant *-l-*, which goes back to the historical or
Proto-Romani demonstrative stem in *-t-* (for the basis of the historical recon-
struction see Matras 2002: 96ff.).

In a number of areas the dialects of the region show different selections
among what can be assumed to have been Early Romani variants (cf. Matras
2002: Ch. 9). Copula forms in *h-* and *s-* seem to have been competing forms in
Early Romani; evidence for this is the presence of two distinct series in some
dialects. The selection of the series in *h-* predominates in western Europe,
especially Germany and France, but it is also found in individual dialects of
southeastern Europe, most notably in Macedonia and Romania. The 2sɢ past

tense and copula person marker *-al* appears to be an archaic form that goes back to pre-European Proto-Romani (see Matras 2002: 143ff.), but it seems to have competed already in Early Romani with the form *-an*, formed by analogy to the second person plural. The distribution of *-al* is largely limited to coherent groups of dialects in central Europe and Scandinavia, but Romania may be said to constitute a transition zone. The alternation of *-s-* and *-h-* in intervocalic position in grammatical paradigms also seems to be old, but selection is geographically contained, with forms in *-h-* being preferred in central Europe and Scandinavia as well as in regions extending along the Adriatic coast to northwestern Greece. Here too, Romania constitutes a transition zone. To the patterns of 'variant selection' we might add two further developments that involve morphological paradigms. The first is the re-structuring of the demonstrative paradigm, based on a selection and/or extension of the inherited Early Romani forms **akava*/*adava* etc. The second is the re-structuring of the verb derivational system of loan verb adaptation markers that was composed in Early Romani of a complex set of Greek-derived markers such as *-in-*, *-iz-* and *-is-*, in combination with inherited valency-marking morphemes such as *-ar-* and *-áv-* (for a discussion see Matras 2002: 128ff.).

Some of the innovations found in the region under consideration are widely distributed across Romani dialects. They include the reduction of final *-s* in the 3SG person marker *-(j)as > -(j)a(h)* as in *dikhljas* 's/he saw' > *dikhlja(h)*, of final *-s* in the second-person singular person marker, as in *aves* 'you come' > *ave(h)*, and of final *-s* in the remoteness marker, as in *sanas* 'you were' > *sana(h)*; and the aspiration of /s/ in positions preceding consonants, as in *leske* 'for him' > *lehke*. Also widespread is the volatility of the phonological process of *v*-prothesis in positions preceding back raised vowels /o, u/: *ušt* 'lip' > *vušt*. Romania is also at the crossroads of a split in the person marker of the 2PL past tense between the more conservative form *-an* for the 2PL, the innovation *-en* (a partial analogy to the third person plural), which is prevalent in southeastern Europe, and the innovation *-e* (a full analogy to the third person plural), which prevails in eastern and parts of central-eastern Europe but is also encountered sporadically in the southeast.

A series of innovations are typically contained within the region under consideration, that is, Romania and some of the adjoining areas in Ukraine, Moldova, Serbia, and northern Bulgaria: the vowel shift in the 1SG person ending *-jom > -em* and in the sequence *-aj- > -ej-* (as in *daj* 'mother' > *dej*, *čhaj* 'girl/daughter'> *čhej*); the loss of affrication in words like *džukel* 'dog' > *žukel* and *čhej* 'girl/daughter'> *šej*; the prothesis of *a-* in words like *šun-* 'to hear'> *ašun-*; the acquisition of adjectival gender inflection on demonstratives, as in *kada* 'this.M' > *kado*; the emergence of the negators *či* and *ni* alongside or replacing the historical negator *na*; the word-specific full patalisation, affrication or

TABLE 2. Features often considered 'diagnostic' of so-called 'dialect branches' that are represented in the region

Feature	Balkan	Balkan-*zis*	Northern Vlax	Southern Vlax	Northern Central	Southern Central
*ṇḍ	r	r	n(d)ř, rn	rn	r, (ndr)	r
2SG past	-an	-an	-an	-an	-al	-al
1SG past	-om	-om	-em	-em	-om	-om
2PL past	-en	-en	-an	-en	-an	-an
3SG copula	(i)si	si	si	hi	hin	hi
Demonstrative	akava	kada/kava	kava/kado	kava	kada/ada	kada/ada
3SG pronoun	ov	ov	vov	vov	ov	jov
Loan-verb marker	-in-	-iz-	-isar-	-isar-	-in-	-in-
'day'	dives	zis	djes, gjes	džive	dives	dive
'with her'	lasa	lasa	lasa	laha	laha	laha
SG.M.OBL definite article	e	e	le	le	le	le
Indicative negator	na	na	či	ni	na	na
'dog'	džukel	džukel	žukel	džukel	džukel	džukel
'hear'	šun-	šun-	ašun-	ašun-	šun-	šun-

other sound substitution in the segments /di, ti, gi/, as in *dives* 'day' >*džes*, *djes*, *gjes*, *zis*; *tikno* 'small' > *cikno*; *tiro* 'your' > *tjiro*, *ciro*, *kjiro*; *gili* 'song' > *djili*, *zili*; and *v*-prothesis in the form of third-person pronouns *ov* 'he' > *vov*.

There are, of course, numerous other developments that are widespread across sub-regions and in individual locations. The selection of the above features for consideration in this study is based firstly on their role as indicative markers of major isoglosses that divide the entire Romani-speaking landscape of Europe into broader zones, which happen to cut right through the Romanian territories; and second on their status as supposedly 'diagnostic' of distinct so-called 'dialect branches' that are represented in Romania and adjoining areas, as can be seen from Table 2.

5. The geographical distribution of features

5.1. *Mixed retention and innovation areas*

The Proto-Romani consonant cluster *ṇḍ can be assumed to have developed in Early Romani into *ndř.[10] Romania represents a conservative region in regard

---

10. For a discussion of 'Proto-Romani' and 'Early Romani' and the method of historical reconstruction for this and other forms see Matras 2002.

nř
r
nr
rn
ndr, ngr
r̩
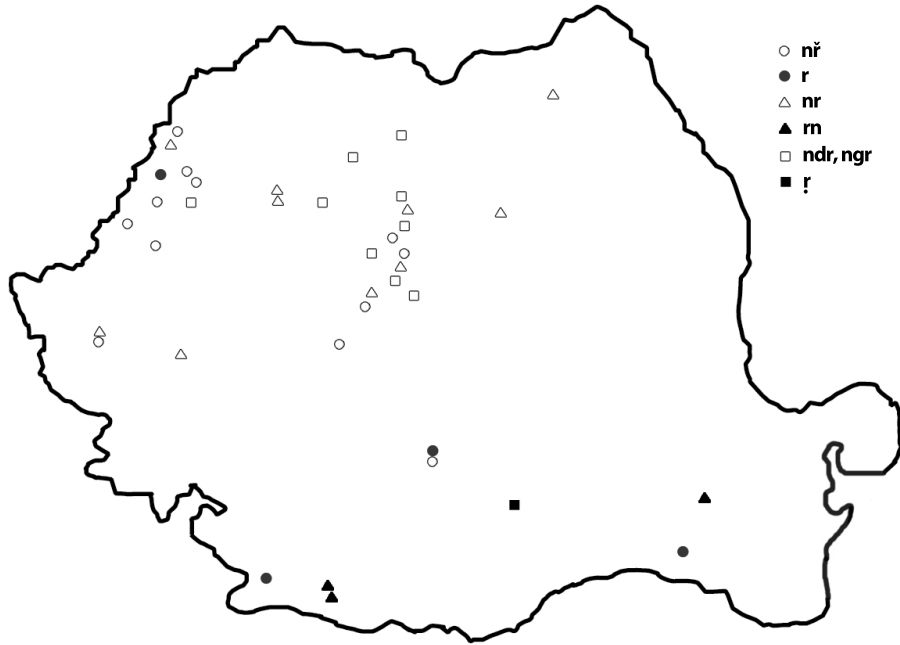
MAP 2. Continuation of *ndř*

to this feature. Map 2 shows that the samples recorded in most locations have preserved a consonant cluster consisting of a trill and a nasal, often with the stop intrusion (*d* or *g*). The reduction of the cluster to *r* is found sporadically, mainly in the south, but in one case (RO-048, Roma Ungrika of Spinus-Oradea) also in the northwest. There is a general division between the more conservative north, which shows predominantly the forms *nř*, *nr*, *ndr* and *ngr*, all representing the original sequence of nasal and trill, some with the original or modified stop intrusion, and the south, which shows the more innovative forms, mainly inversed *rn* and simplified *r*, and in one case (RO-021, Spoitorja of Bolentin Vale) a simplification of the original cluster to retroflex *r̩*.

The vowel fronting in the 1SG person marker *-(j)om* > *-em* and in positions preceding palatals (or diphthongs) as in *čhaj* 'girl, daughter' > *čhej* is often considered a diagnostic feature of the Vlax Romani dialects (cf. Boretzky 2003). Maps 3 and 4 show that the distributions of these two developments are in fact quite distinct.

The first, in the 1SG marker (Map 3), is triggered presumably by the historical palatal glide that preceded the vowel segment and is thus merely a 'pseudo-umlaut', while the second is a 'genuine' umlaut triggered by the following palatal glide. In the person marker, the geographical distribution of the change appears to be random, but shows a tendency to correlate with the group names Kəldərarja/Kelderari/Kalderaša (samples RO-008, RO-052, RO-065, respect-

om
em
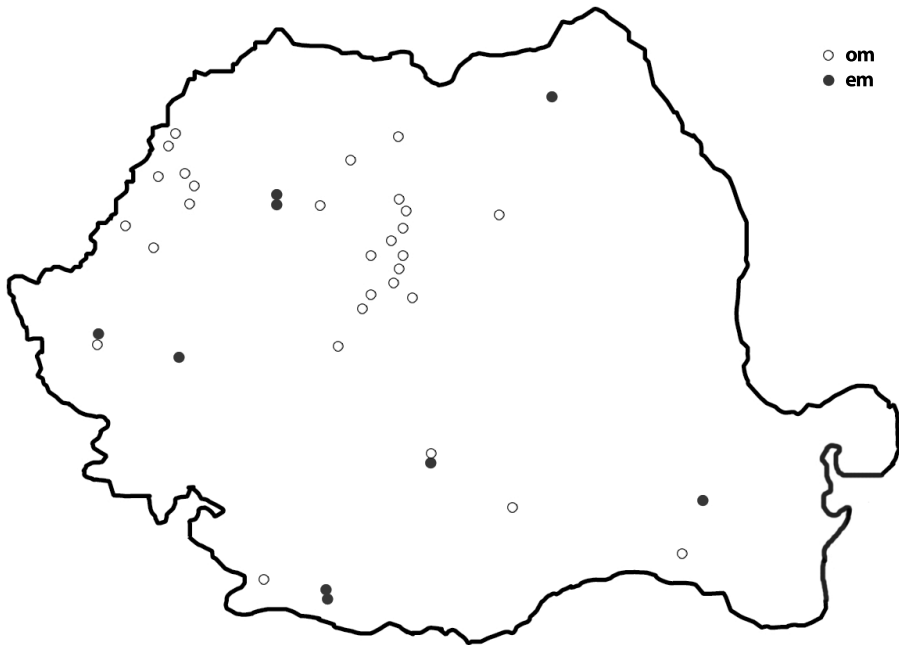
Map 3. Vowel umlaut in *-om > -em* (1sg past/copula)

ively) as well as others (ro-013 Piculesči Roma, ro-015 Kurturare, ro-020 Ursarja, ro-064 Kangljari, and in one case, ro-054, even Ungrika Rom). For the Kelderash group and in all likelihood also the Kurturare, we might on this basis postulate an historical dispersion of what were once coherent groups. By contrast, the umlaut *-aj > -ej* (Map 4) is found predominantly in the north, but it is also present in the speech of the Ursarja and Kərəmidarja of Sadova (ro-020 and ro-074 respectively) and of the Kangljari of Ţăndărei (ro-064) in the south, while the north shows some small retention zones of the con-servative form. We might therefore describe this pattern as an innovation with sporadic retention in the north, contrasting with retention with sporadic inno-vation in the south. We also find sporadic retention of the 2sg past tense and copula person marker *-al* (not depicted on the maps) among the Gabor, Teglari and Rom Mătase of Mureş country (samples ro-009, ro-056 and ro-053) and among the Gabor of Timişoara (sample ro-011)

   Mixed diffusion patterns of a somewhat comparable nature can be recog-nised for other developments, too: the reduction of final *-s* in the 3sg past (*dikhljas* 's/he saw' > *dikhlja(h)*) (Map 5) appears with a high density in the north but is also present, albeit less prominently, in the south. The appearance of prothetic *v-* in *ušt* 'lip' > *vušt* (Map 6) is by contrast sporadic in the north and by and large absent in the south (with the exception of sample ro-021, Spoitorja from Bolentin Vale).

aj
ej
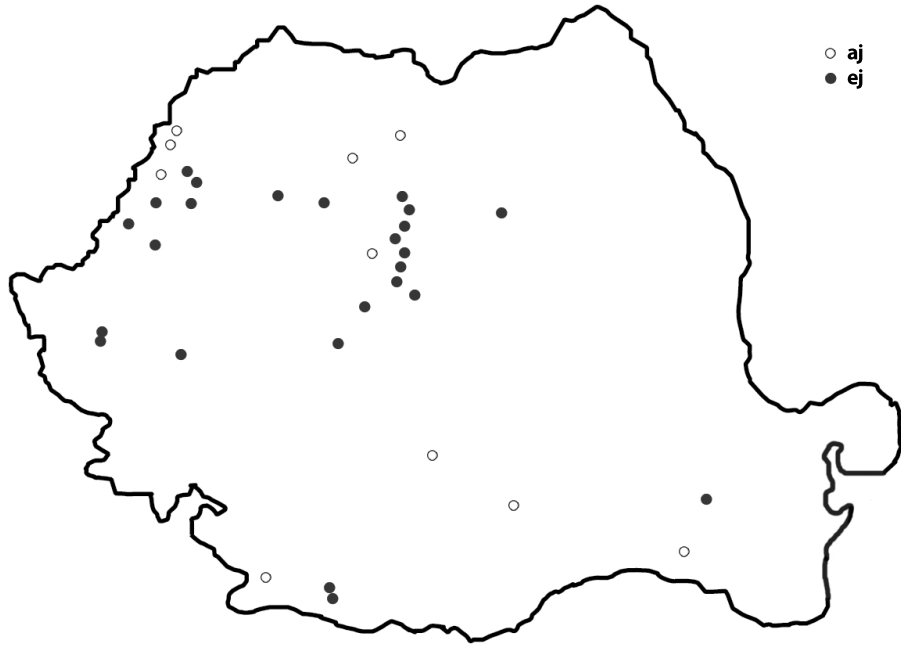
MAP 4. Vowel umlaut in *čhaj > šej, čhej* 'girl'

○ **-as**
● **-a(h)**
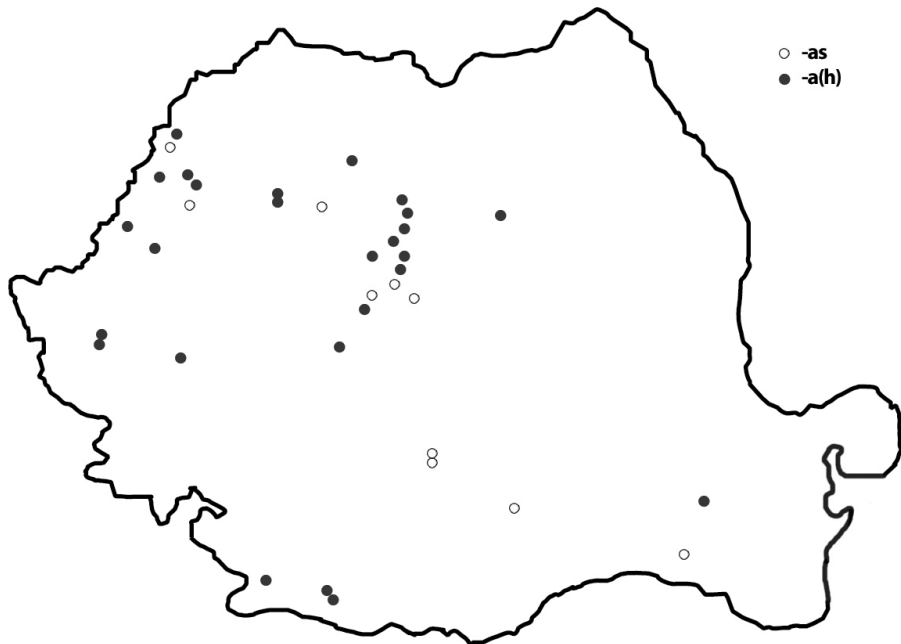
MAP 5. Loss of *-s* in 3sg past: *dikhljas > dikhlja*
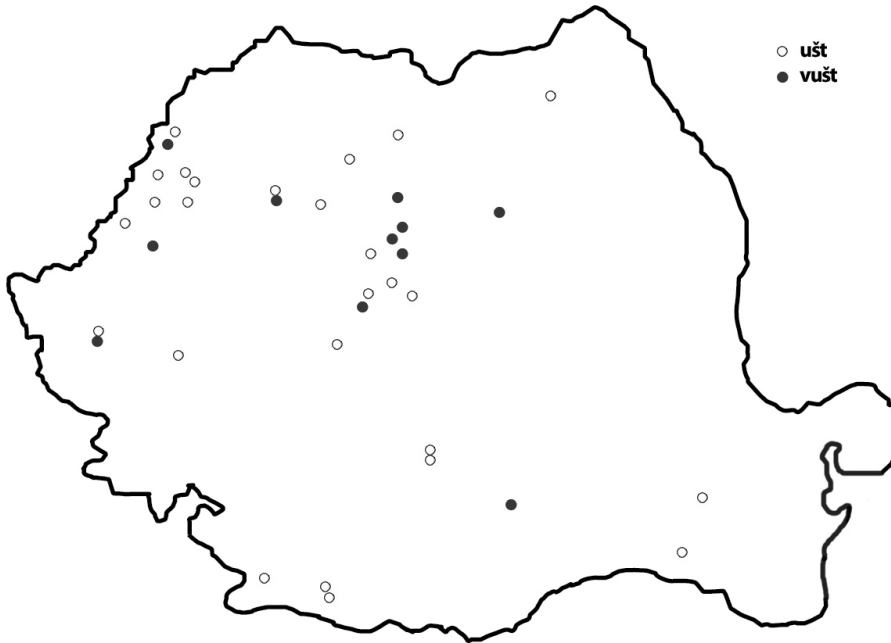
Map 6. Prothetic *v-* in *ušt* > *vušt* 'lip'

A pattern of dymanic local developments leading to diversity across the area under consideration is represented by the various degrees of palatalization of *tiro* 'your' > *kjiro, čiro* (often also in shortened form *to, čo*) (Map 7). We find both the conservative form in *t-* and the lightly palatalised form in *kj-* through-out the country, while the more advanced affrication to *č-* is attested only in the north, albeit sporadically. The pattern resembles that of the changes to the initial consonant in *gili* 'song' > *zili, džili* (not depicted on the maps). Here, the conservative form in *g-* and the shift to a fricative in *z-* are both widespread across the country, while the affricate form *dž-* is only found, sporadically, in the north, often in the same samples that show *čiro* 'your'.

The appearance of a prothetic vowel in *šun-* 'to hear' > *ašun-* (Map 8) may be considered typical of the diffusion centre that is based in Romania, but as we can see it actually represents a case of widespread innovation with isolated, sporadic retention areas. A similar pattern is found for the de-affrication of the initial consonant in *džukel* 'dog' > *žukel* (Map 9).
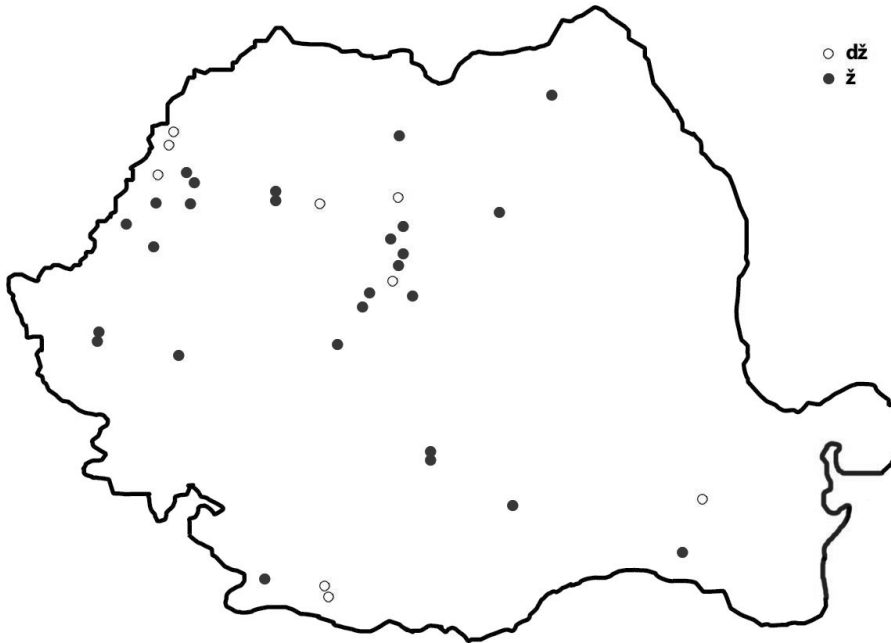
Maps 2–9 thus illustrate the dynamics of a transition spread zone, one in which innovation is often widespread but does not reach all the varieties of the region, whereas in some cases innovation appears to be incipient and so still contained.

tiro/tro (to)
čiro (čo)
kiro, kjiro
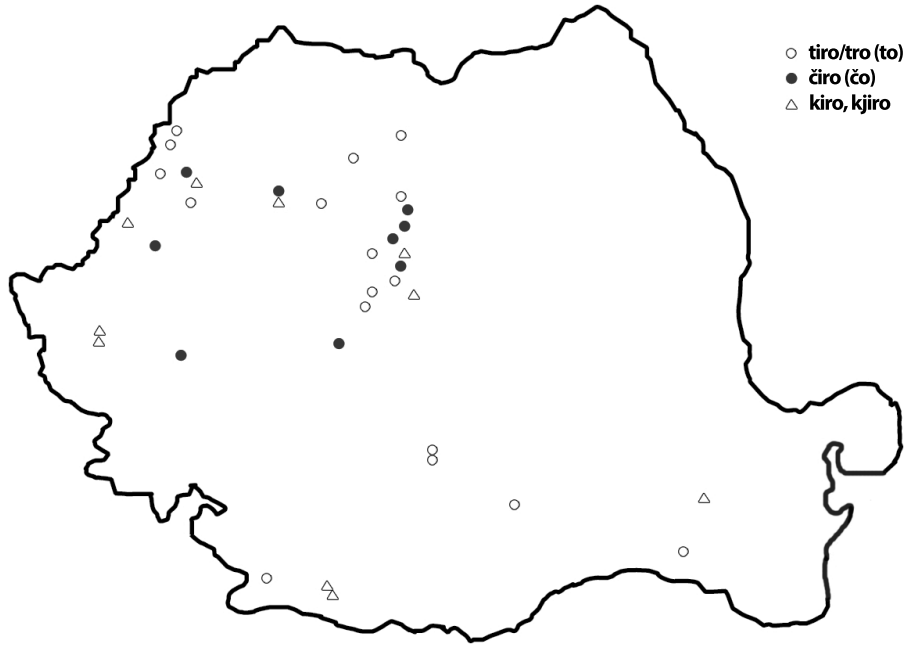
MAP 7. Palatalisation in *tiro* (*to*) > *čiro* (*čo*) 'your'

šun
**ašun**
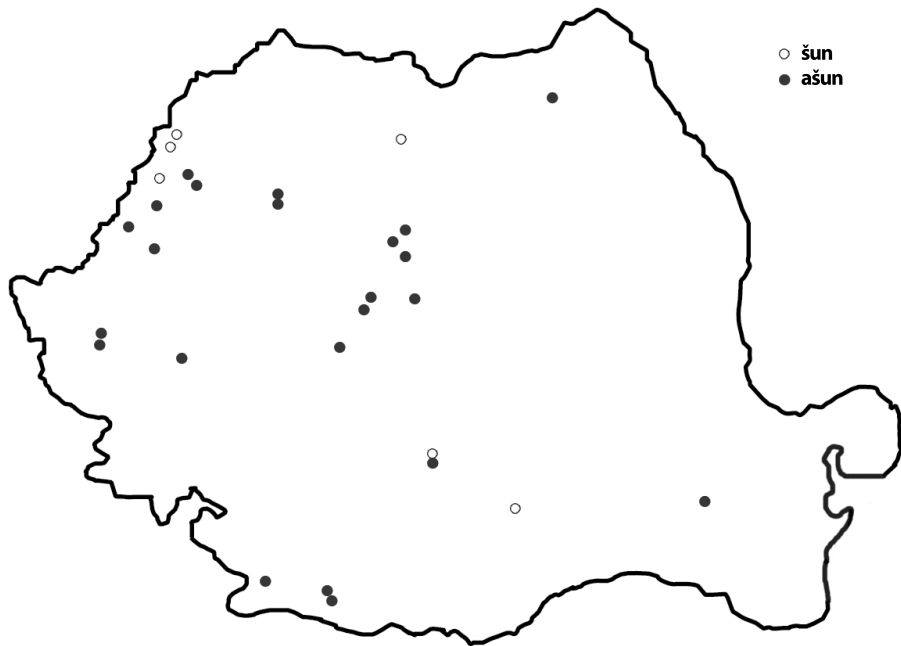
MAP 8. Prothetic *a* in *šun* > *ašun* 'to hear'

MAP 9. Consonant de-affrication in *džan > žan* 'to know' and *džukel > žukel* 'dog'
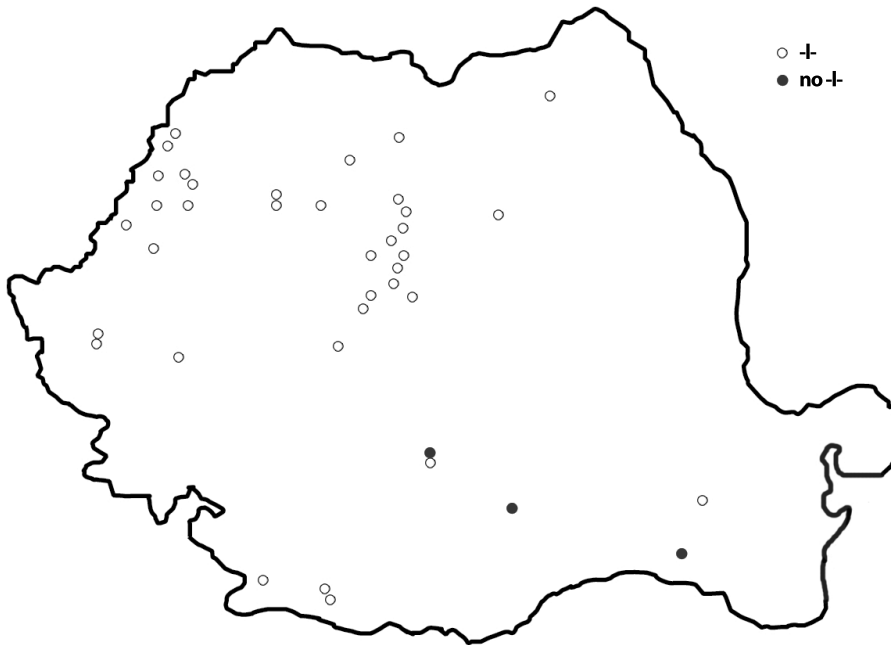
5.2. *The North–South divide*

The mixed pattern prevails in a series of additional developments, yet some allow us to recognise somewhat clearer regional tendencies. While Romania is often associated with Vlax forms, the dominant form of the indicative negator is in fact the conservative *na* (Map 10). The form *či* that is typically associated with Northern Vlax (cf. Boretzky 2003) is a widespread innovation that is found primarily in the centre of the country, albeit not consistently and often alongside *na*, while *ni* is limited to the south, appearing in a coherent zone in the southwest corner of the country as well as in the Kangljari dialect of Ţăndărei (RO-064). The changes to the initial and medial sounds in *dives* 'day' > *des*, *djes*, *džes*, *gjes*, *zis* and more (Map 11) offer a further insight into regional sub-divisions: The shortened forms *des*, *djes*, *džes*, *gjes* and variants are spread across the country, while the more conservative *dives/djives* is retained only in a few samples from the Bihor county area in the northwest (with self-appellations Ungrika Rom and Lovari), and the shift to the initial fricative *zis*, *zi* (considered diagnostic of the so-called Balkan-*zis* group, or South Balkan II in the terminology favoured by Boretzky 2000) is confined to the Spoitorja samples from the southeast. The samples from the very same sub-group, the Spoitorja of the southeast, are also the only samples that do not preserve the original *-l-* in the oblique definite article (Map 12).

na
či
na/či
ni

MAP 10. Indicative negation marker
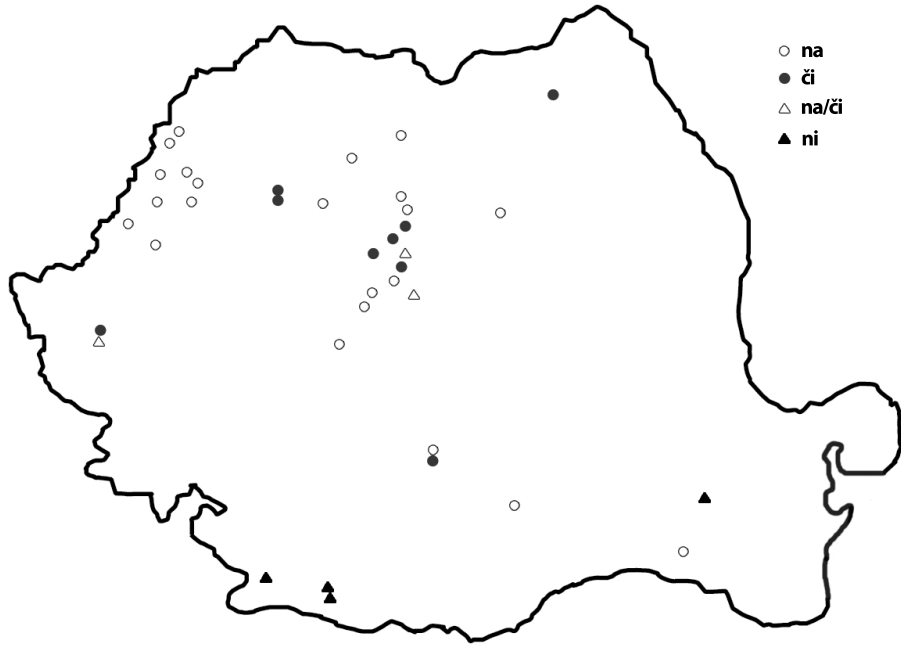
dives, djives
des, djes, dže, džej
gjes, ges
zis, zi
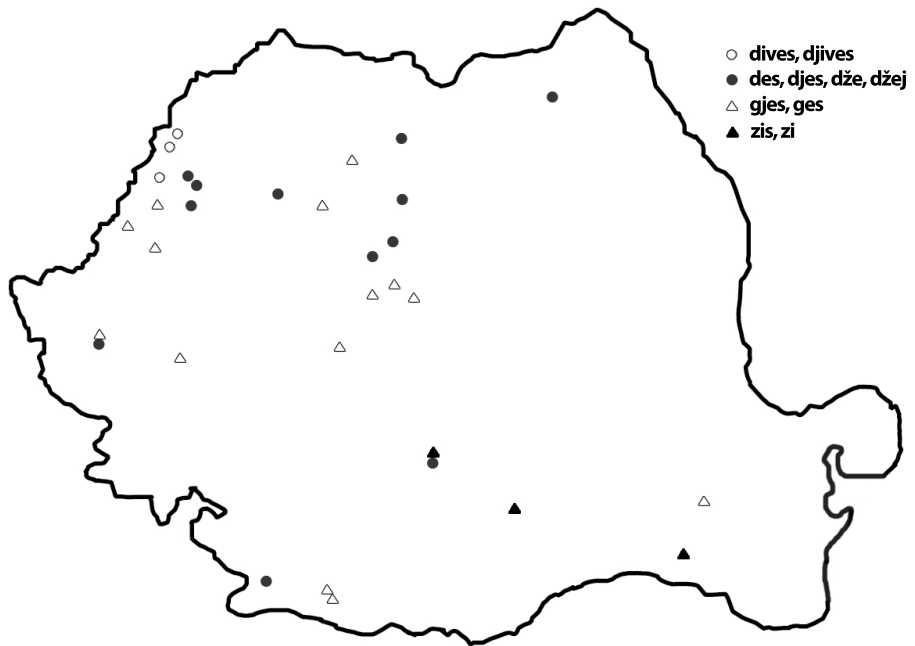
MAP 11. Sound reduction and sound substitution in *dives* > *djes*, *gjes* > *zis*, etc. 'day'

-l-
no -l-

Map 12. Presence of *l-* in oblique definite article (<**ol-*)

The geographical picture of a north–south divide becomes even clearer when we consider a series of developments involving both phonological and morphological innovations, and in some cases historical 'option selection' (from among inherited Early Romani variants; see discussion above). The north-central region around Mureş county is the centre for the loss of the final -*s* in the person marker of the 2sɢ present *aves* 'you come' > *ave(h)* (Map 13), though even within this area the development is not consistent. Its spread to the variety of the Gabor of Timişoara (ro-011) can be attributed to the close links among the Gabor group, whose main areas of settlement are Mureş county and Timişoara. Contained within the same area around Mureş county are two further developments involving the reduction of -*s*-, namely the aspiration of -*s*- in pre-consonantal position as in *leske* 'for him' > *lehke* (Map 14), and the loss of -*s* in the remoteness marker as in *sanas* 'you were' > *sana(h)* (Map 15). Once again we witness the spread to the variety of the Gabor of Timişoara (ro-011).

The de-affrication of the initial consonant in *čhej* 'girl' > *šej* (Map 16) also seems to have its centre in Mureş county in the centre-north, extending to Timiş county in the west, and here too the south remains unaffected by the development. Three additional developments form very similar geographical distribution patterns: The selection of copula stems in *h-* (for all persons) (Map 17) appears sporadically in the area roughly between Mureş and Bihor

-es
-e(h)
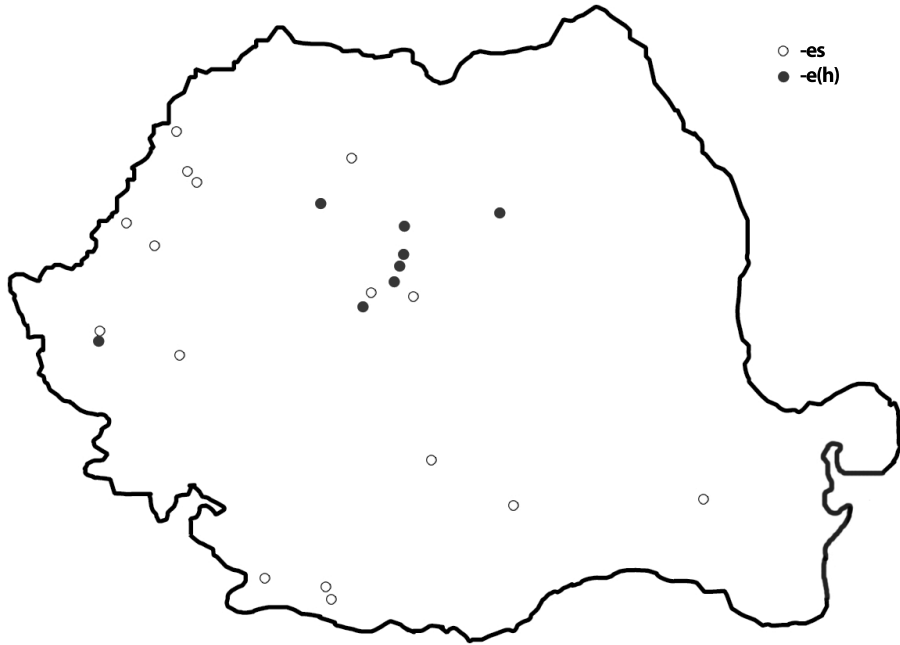
Map 13. Loss of -*s* in 2sg present: *aves* > *ave*(*h*)
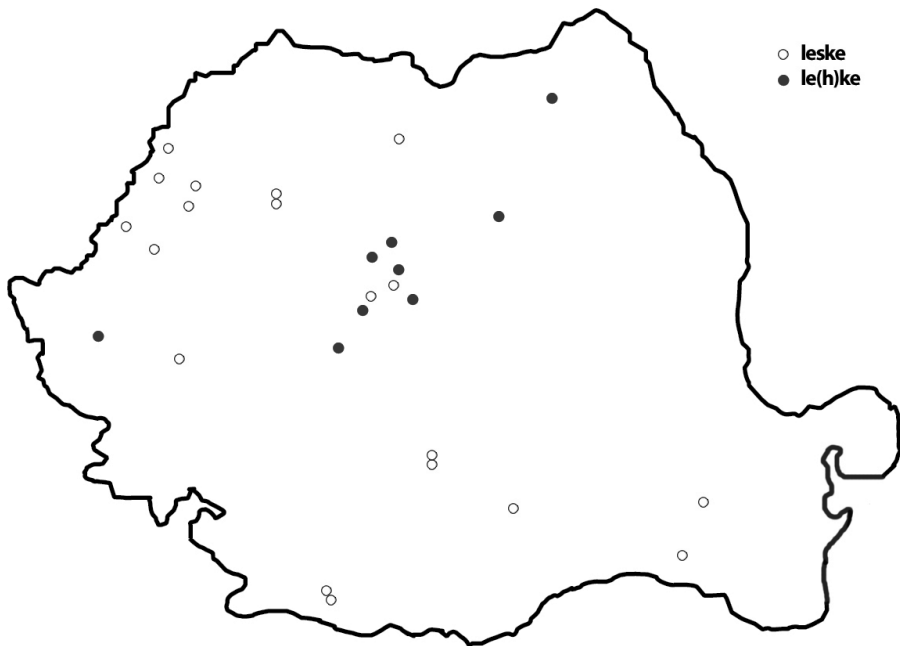
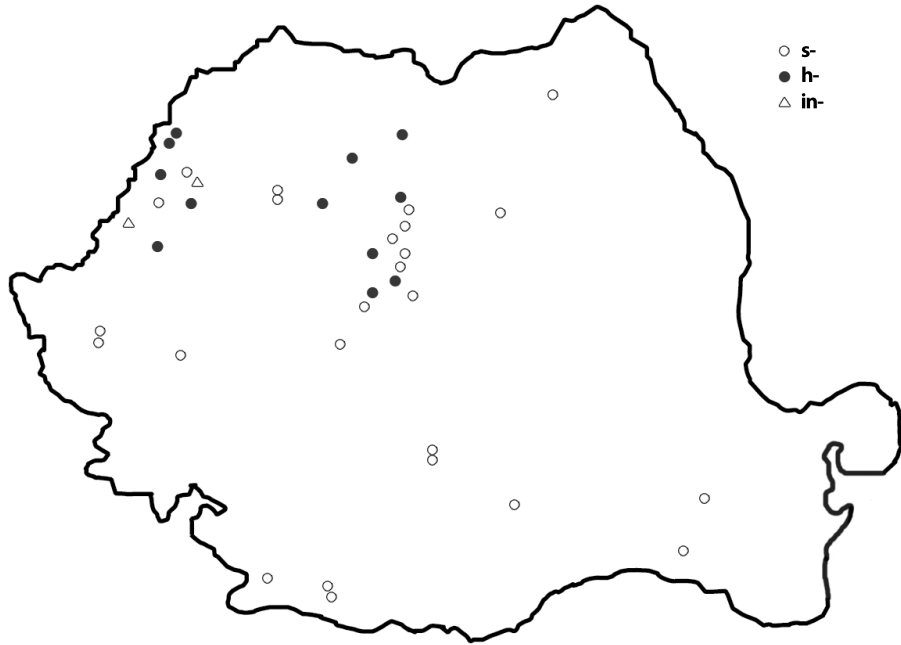○ **leske**
● **le(h)ke**

Map 14. Loss of -*s* in pre-consonental position: *leske* > *le*(*h*)*ke* 'for him'
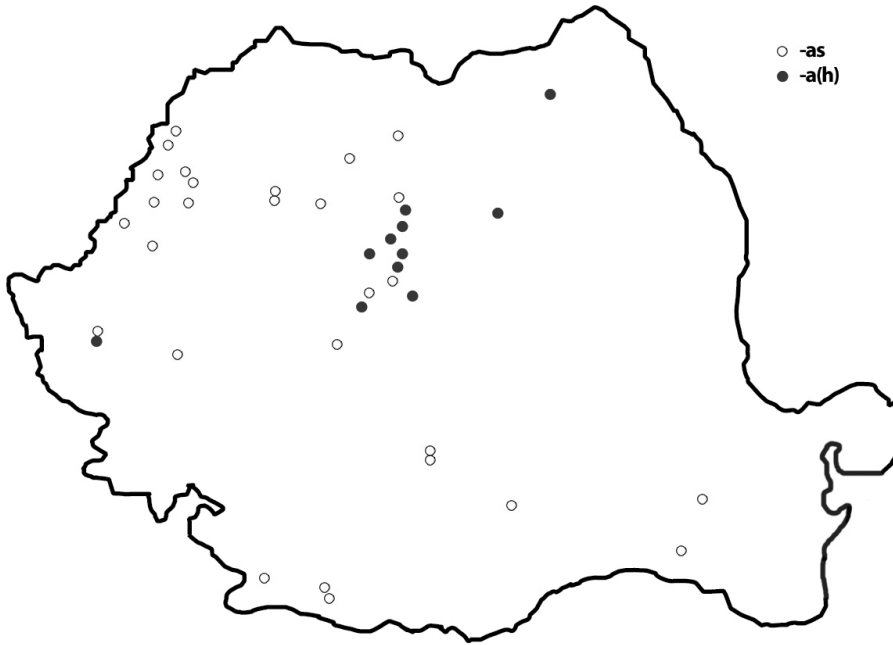
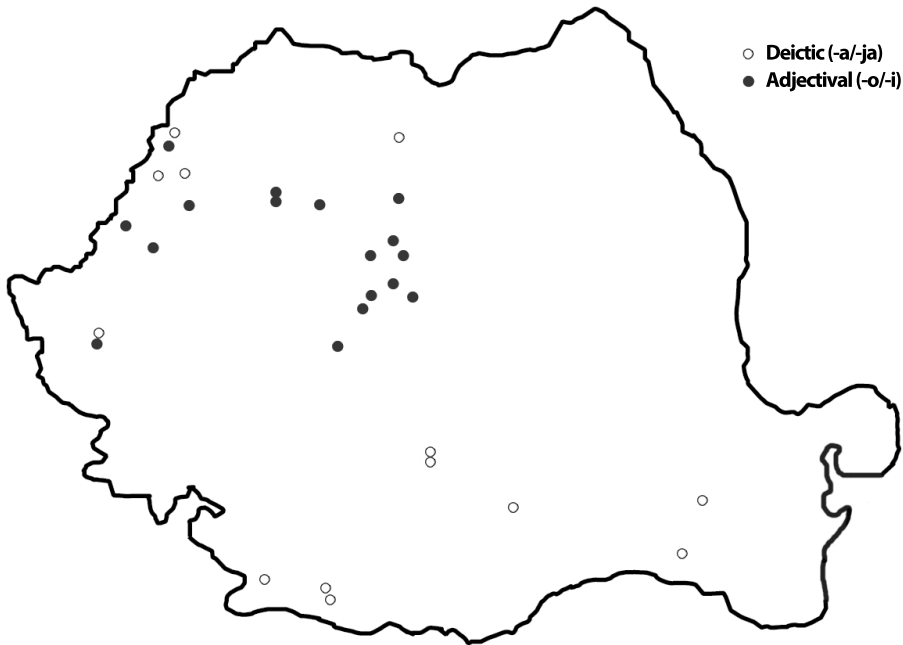**Map 15.** Loss of is *in the remoteness marker: sanas > sana(h)* etc.

čh
š
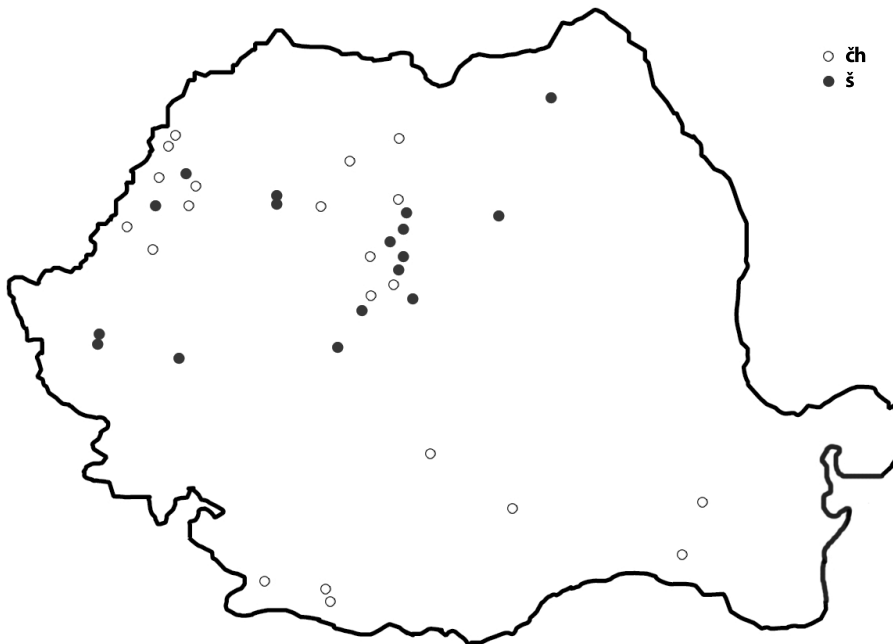
Map 16. Consonant de-affrication in *čhej > šej* 'girl'

s-
h-
in-

Map 17. Copula stem
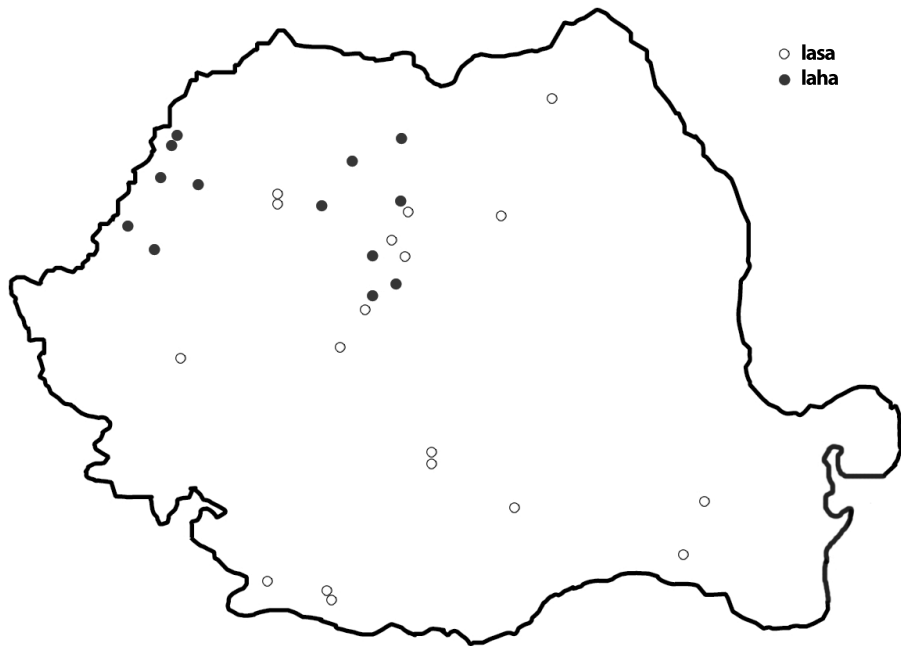
lasa
laha

Map 18. Aspiration of intervocalic *-s*: *lasa* > *laha* 'with her'

Deictic (-a/-ja)
Adjectival (-o/-i)

MAP 19. Gender inflection in demonstratives

counties, matching rather closely the selection of *-h-* in intervocalic positions in grammatical paradigms: *lasa* 'with her' > *laha* (Map 18). Note that the two developments tend to stand in hierarchical correlation to one another in Romani in general: dialects that select an *h*-copula almost always select intervocalic *-h-* in paradigms, but not vice versa (cf. Matras 2002: 68–70).

A completely separate development is the extension of adjectival gender inflection to demonstratives (Map 19). The Early Romani pattern, which continues in the majority of the dialects, has specifically deictic masculine forms in *-(v)a* and feminine forms in *-ja* (e.g. *adava/adaja*). The innovation centred in this area shows an analogy to the gender inflection of adjectives, with demonstratives acquiring the inflectional endings *-o/-i* (*ado/adi*, *kado/kadi*), while plural number inflection may pattern with either the older, deictic system or the adjectival system.

A similar overall pattern is represented by the distribution of word forms for 'tomorrow' (Map 20). The northern zone appears by and large coherent: Here the preference is for *tese/tehe* and variants, with the split between forms in *-s-* and those in *-h-* shadowing the overall volatility of intervocalic *s/h* alternation in the region. The form *tehara* is widespread across the country but infiltrates the north rather selectively, being confined largely to samples with self-appellation Kalderaša, Kelderari, Kurturare, and Piculesči Roma (RO-065 in Timişoara, RO-052 in Suceava, RO-015 and RO-016 in Huedin, respectively)

tese, tejsa
**tehe, the**
tehara
ajnara
rano
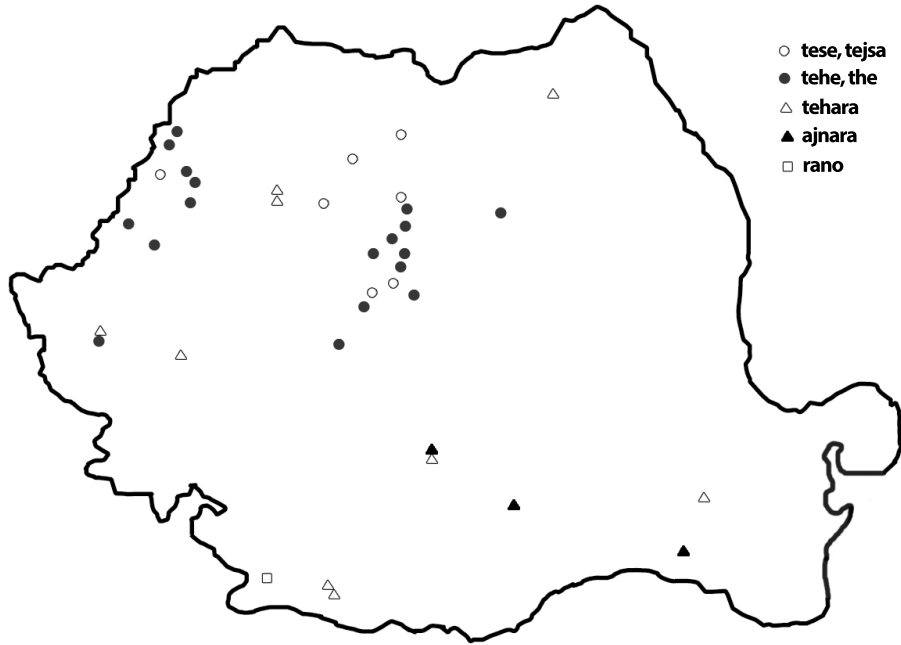
MAP 20. 'Tomorrow'

as well as Ungrika Roma (RO-054 of Caransebeş). The three Spoitorja samples in the south (RO-006, RO-021, RO-016), on the other hand, show consistently their distinctive form *ajnara*.

5.3. *Regional spread zones*

Above we have already witnessed how a number of developments are confined to geographical sub-divisions within the broader north–south divide, the cluster of samples in the extreme north-western corner of the map (Bihor county) being a case in point, another being the north-central area around Mureş county. The final batch of features that I would like to discuss involves the formation of morphological paradigms. In Matras 2002 (Ch. 9) I argued that the re-structuring of morphological paradigms in Romani tends to result in rather coherent geographical spread zones. In fact, it is primarily the isoglosses that separate morphological paradigm formation (e.g. demonstratives, person markers, and loan verb adaptation markers) that match the impressionistic so-called 'consensus' classification of Romani dialect groups most closely (cf. also Matras 2005). The explanation that can be offered for this finding is that paradigm re-structuring is a complex and less predictable process, prone to a variety of factors such as levelling, analogy, reduplication and word-combination, grammaticalisation and more. The fact that speaker communities undergo such a complex innovation process in a shared and coherent manner implies

adauka
kadava
**kada, kado**
ada, ado
kava
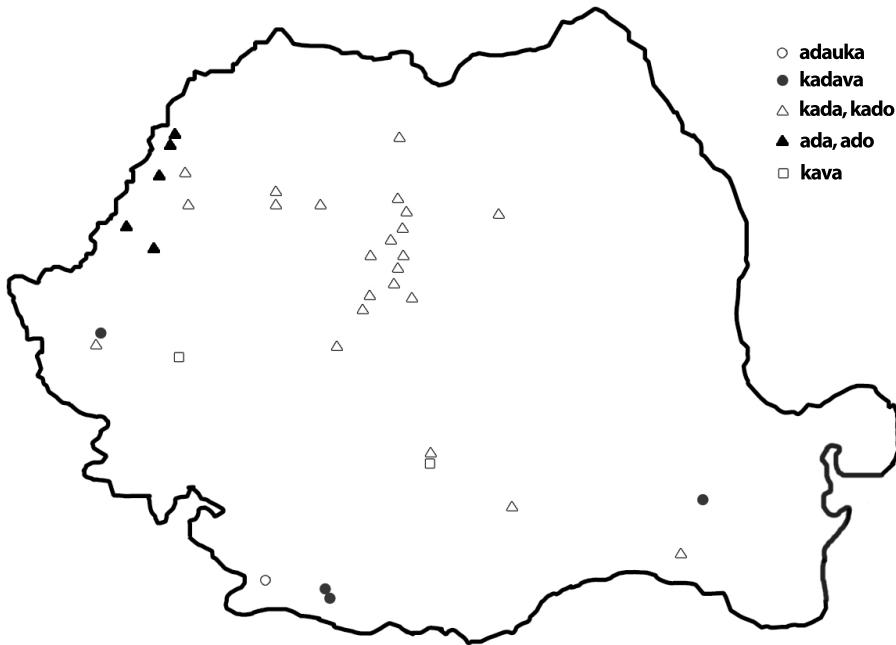
MAP 21. Demonstrative stems

prolonged and intense contacts and is therefore an indicator of historical social coherence (in the form of a tight network of contacts) within a population group. Our case study for Romania is a nice test-case for similar principles, all the more so since some of the major isoglosses that separate the Romani linguistic landscape at a pan-European level cut right through Romanian territory.

Romani dialects inherited a complex, four-term system of demonstratives from Early Romani (*adava/akava/odova/okova*). The original system tends to be preserved in the geographical periphery (southern Balkan, southern Italy, Wales), while in the centre the forms are simplified but also re-structured to preserve, by and large, a four-term semantic differentiation (cf. Matras 2002: 103ff.). In Romania, the most widespread demonstrative stem is *kad-* (Map 21). The forms *kadava* and *kava* have a sporadic presence and seem to be randomly distributed across both regions and groups, though they are both absent from the north. The north itself is split, with its extreme western zone around the counties of Bihor and Arad showing forms in *ad-* (typically associated with the so-called 'Central' dialects of Romani).

A smooth north–south split is found for the distribution of the 3SG pronoun form (Map 22), with the north consistently showing *v*-prothesis to *vo(v)*, which is consistently absent in the south. This overrides any group affiliation, though two of the three Spoirorja dialects in the south show a substitution of the pro-

ov, ou
vo(v)
oda
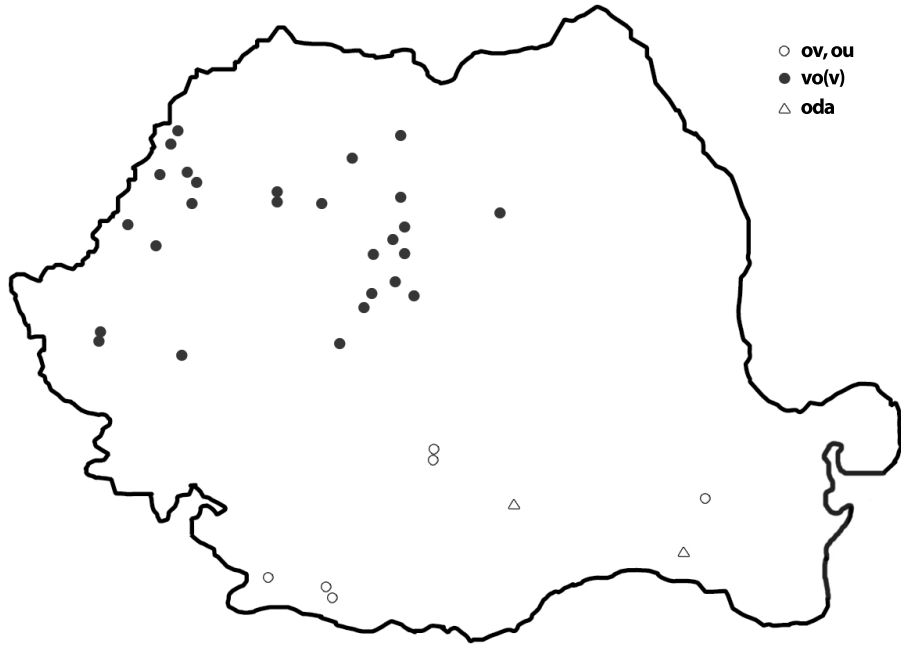
MAP 22. Third person singular masculine pronoun

kici
sodi, sode
kibor, kabor
kazom, kozom
kicom
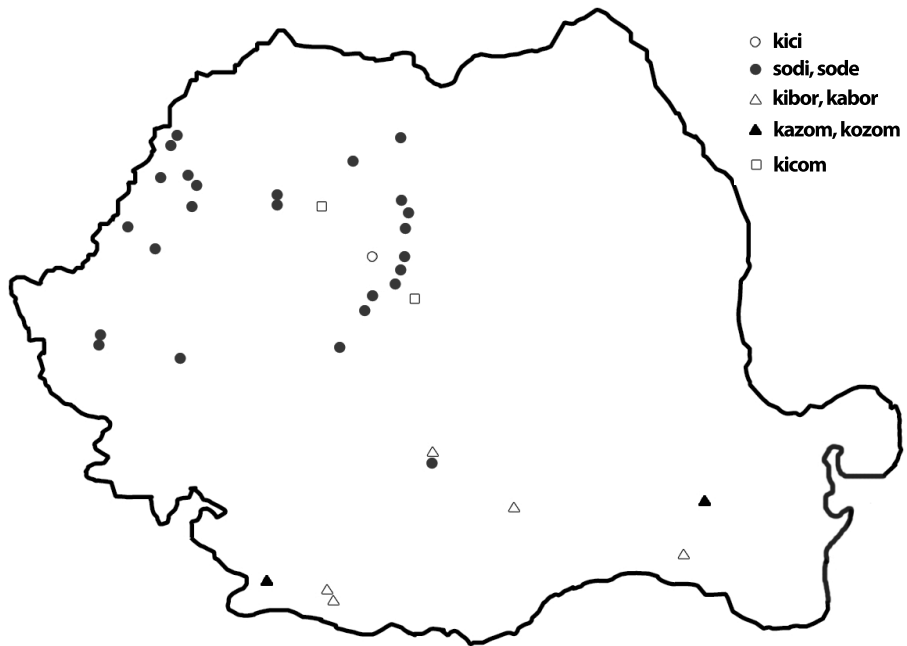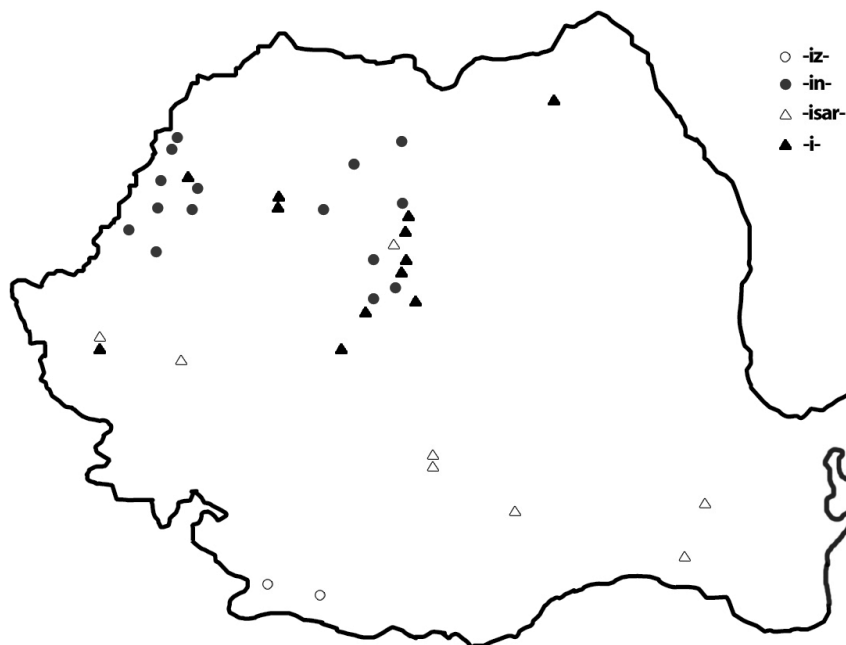
MAP 23. 'How much'

MAP 24. Loan-verb integration marker

noun through an historical demonstrative form *oda*. A north–south split is also prevalent in the formation of the interrogative 'how much' (Map 23), which is prone to considerable variation in Romani. The north has mainly *sode/sodi*, with isolated occurrences of *kicom* and *kici*. While *sode* is a regional innovation found almost exclusively in this particular zone, the forms in *kic-* can be associated with *kiti*, which is the principal form found across the Romani dialects of central and eastern Europe (i.e. north of the Great Divide). The south of Romania, by contrast, has *kibor/kabor* and *kazom/kozom*. This matches the forms found in the southern area of Europe (Bulgaria, Macedonia, Serbia, Croatia, Greece, southern Italy, and the Crimean peninsula). Note that here, as with some other features, the Kəldərarja dialect of Piteşti in the south (RO-008) aligns itself with the other Kelderash samples.

The system of loan verb integration markers is perhaps the most complex morphological paradigm in Romani (see discussion in section 4, above; cf. Matras 2002: 128ff.). Here too we find a north–south split, albeit with the samples from Timiş county taking an intermediate position (Map 24). The north has typically *-in-* (forming part of a continuum with the dialects of central and eastern Europe) as well as the reduced form *-i-*. The form *-isar-* that is usually associated with the Vlax dialects is prevalent in the south, including the samples whose speakers identify as Spoitorja, with an isolated occurrence among the Gabor in Mureş county (RO-009), while *-iz-*, typically associated with the

-an
-en
-e
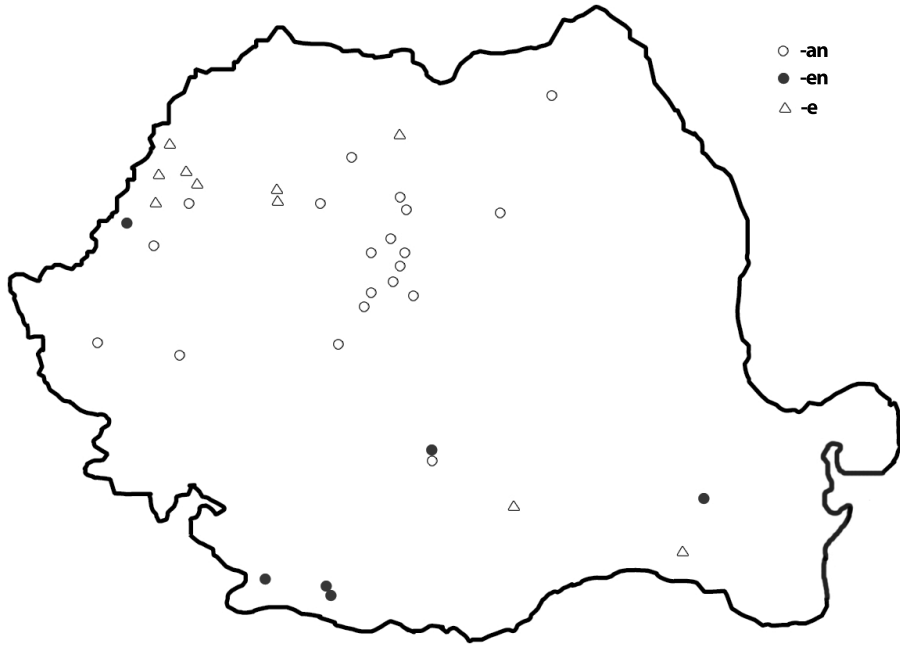
Map 25. Second person past

South Balkan II or *zis*-group, is found exclusively in the two Ursarja samples (RO-004 and RO-020) in the southwestern corner of the country (Dolj county).

Finally, the forms of the 2PL marker in past tense verbs confirms some of the regional divisions seen earlier (Map 25). The conservative form *-an* prevails in the north (and is also found, like some other northern forms, in the Kǝldǝrarja dialect of Piteşti in the south, RO-008). A northwestern zone in Bihor county sets itself apart from the north, however, showing an analogy to the 3PL in *-e*; the same development is found again in two of the Spoirorja samples in the southeast. Otherwise, the south generally shows *-en*, a partial analogy to the 3PL, and the form that prevails across the Romani dialects of the southern Balkans.

6. Group affiliation and dialect

On the basis of the discussion in the previous section, especially the consideration of morphological paradigms, we might generalise and say that Romania shows a rather consistent split in its Romani dialect landscapes: a southern zone aligns itself primarily with the dialect features of the southern Balkans (Bulgaria, Serbia, Macdeonia). A small northwestern zone centred in Bihor and Arad counties aligns itself with the dialects of central Europe (Hungary, Slovakia, Poland). A northern and often north-central zone shows a density

of particular developments, such as the interrogative *sode* 'how much', the reduced loan verb marker *-i-*, adjectival inflection on demonstratives (*ado/ kado*) and *v*-prothesis on the 3SG pronoun *vov*. These developments are neither typical of Romania as a whole, nor are they confined to dialects that show the 'typical' profile of either Vlax or Northern Vlax varieties or whose speakers necessarily identify as Kelderash, Čurar, Kurturar or other groups names that are associated with 'Vlax'; in fact, the same features are often found both among recent group denominations like 'Gabor' and among established self-ascriptions such as 'Hungarian Rom' (Ungrika Roma or Romungro).

To what extent can we nevertheless associate dialect features with self-ascription – independently of geographic location? Table 3 (overleaf) divides the majority of the samples included in this study into six major groups, leaving out some of the isolated occurrences of reported endonyms in just single samples. Group 1 lumps together the Romungro (Hungarian Rom) with the synonymous Ungrika Rom and Roma Ungrika. Group 2 includes the Spoitorja and Group 3 the Ursarja. Group 4 includes several sub-divisions, which are justified partly on the basis of self-reports by the speakers about their affinity with other groups, partly on the basis of reports in the literature on connections among so-called Vlax groups. Group 5 includes both the self-ascribed Gabor and the Šušuwaje/Žuržuvaje, which according to speaker reports as well as ethnographic descriptions (cf. Jacobs 2012) are all closely affiliated groups.

Considerable coherence is found among the majority of the dialects of Group 1 in Bihor, Cluj and Mureş counties who refer to themselves as Romungro or Rom Ungrika. By contrast, those who refer to themselves as Ungrika Rom in Bihor (sample RO-063) show a number of distinct features, as do the Ungrika Rom of Miriş county (sample RO-054). The typical features of the Romungro cluster are the retention of person marker 1SG *-om* and of the affricates in *čhaj* 'girl' and *džukel* 'dog', absence of palatalisation in *tiro* 'your' and *gili* 'song', loss of final *-s* in the remoteness marker *sana* 'you were', selection of *-h-* in intervocalic grammatical morphemes like *laha* 'with her' and of *h-* in the copula paradigm, retention of *-l-* in the oblique definite article, *v*-prothesis in the 3SG pronoun *vov*, retention of *na* as indicative negation marker, selection of the loan verb integration marker *-in-*, and a tendency toward demonstratives in *ad-*. Many of these features are in line with those associated with the Romungro (Southern Central) dialects of Hungary and Slovakia, but our samples differ from those in showing prosthesis in the 3SG pronoun *vov* and the absence of the 2SG past tense and copula person marker *-al* (using *-an* instead).

Strong coherence is also found among the three samples of Group 2 (Spoitorja), which are located in different counties in the south. The typical features of the Spoitori cluster are the retention of the original vowel in *čhaj* 'girl' and in the person marker 1SG *-om*, retention of the affricate in *čhaj* 'girl'

TABLE 3. Dialect groups based on speakers' self-designations

Group number	Group self-designation	Sample number	Location
1	Romungro	RO-001	Cluj
	Romungro	RO-072	Glodeni
	Romungro	RO-059	Bahnea
	Roma Ungrika	RO-048	Spinus-Oradea
	Ungrika Rom	RO-054	Caransebeş, Timiş
	Ungrika Rom	RO-063	Săcuieni
2	Spoitorja	RO-006	Piteşti
	Spoitorja	RO-021	Bolentin Vale
	Spoitorja	RO-016	Călăraşi
3	Ursarja	RO-004	Maglavit
	Ursarja	RO-020	Sadova
4a	Kəldərarja	RO-008	Piteşti
	Kelderari	RO-052	Vereşti-Hancea, Suceava
	Kalderaša	RO-065	Timişoara
4b	Čurari	RO-071	Uileacu de Cris
	Čurarja	RO-051	Velt
4c	Kərəmidarja	RO-025	Deaj
	Kərəmidari	RO-050	Tinca
	Kərəmidarja	RO-074	Sadova
4d	Vlaši Roma	RO-058	Senereus
4e	Kurturare	RO-015	Huedin
4f	Piculesči Roma	RO-013	Huedin
5	Gabor	RO-009	Valureni
	Gabor	RO-010	Crăciuneşti
	Gabor	RO-031	Crăciuneşti
	Gabor	RO-011	Timişoara
	Gabor	RO-019	Corneşti
	Gabor	RO-026	Budiu Mic
	Šušuwaje	RO-012	Târgu Mureş
	Žuržuvaje	RO-036	Sangeorgiu de Mureş
6	Rom Mătase	RO-049	Dej
	Rom Mătase	RO-053	Albeşti, Mureş

contrasting with its loss in *žukel* 'dog', absence of palatalisation in *tiro* 'your', emergence of an initial fricative in *zis* 'day', retention of *-s-* in all positions, loss of *-l-* in the oblique definite article, a tendency toward replacement of the 3SG pronoun by the demonstrative *oda*, retention of *na* as indicative negation marker, use of demonstratives in *kada* (with deictic inflection), selection of the loan verb integration marker *-isar-*, and the use of *ajnara* 'tomorrow'.

Contrasting with the established profile of the southern Balkan *zis*-dialects (South Balkan II) we note especially the use of *-isar-* rather than *-iz-* with loan verbs.

The two samples of Group 3 (Ursarja) form a tight geographical cluster in Dolj county. They both show retention of the affricate in *čhaj* 'girl', retention of *-s-* in all but the 3SG past tense marker *dikhlja* 's/he saw', selection of the loan verb integration marker *-iz-*, retention of *-l-* in the oblique definite article, absence of *v*-prothesis in *ušt* 'lip' and in the 3SG pronoun *ov*, presence of *a*-prothesis in *ašun-* 'to hear', use of *ni* as indicative negation marker, and absence of adjectival inflection on demonstratives. Many of these features, however, are shared with the entire southern or at least the southwestern area. On the other hand, there are quite a few features that set the two samples from Maglavit (RO-004) and Sadvoa (RO-020) apart: the Maglavit sample has the demonstrative *adauka* (Sadova *kadava*), a single consonant in *maro* 'bread' (Sadova *marno*), 1SG *-om* (Sadova *-em*), original vowel in *čhaj* 'girl' (Sadova *čhej*), absence of palatal in *tiro* 'your' (Sadova *kjiro*), and use of *kazom* 'how much' (Sadova *kibor*) and *rano* 'tomorrow' (Sadova *tehara*). On this basis it seems difficult to postulate a structural type based on the group name 'Ursari'.

Distinctive of Group 4a (Kelderash or 'cauldron makers') is the vowel shift in the person marker 1SG *-em*, a tendency toward retention of *-s-* in all positions, use of *či* for indicative negation, a tendency toward selection of the loan verb integration marker *-isar-*, a tendency to use *sode* 'how much', and use of *tehara* 'tomorrow'. A number of features are also shared with groups 4b, 4c, and 4d, as well as with the two samples from Huedin (4e and 4f): de-affrication in *šej* 'girl' and *žukel* 'dog', retention of *-l-* in the oblique definite article, marker, *v*-prothesis in the 3SG pronoun *vov* but absence of *v*-prothesis in *ušt* 'lip', and *a*-prothesis in *ašun-* 'to hear'. These are very much in line with the features that are typically associated with Northern Vlax Romani, but the sample data show that the self-ascribed Kelderash group is indeed more coherent linguistically than the cluster of 'Vlax' dialects that includes the Čurari ('sieve makers'), Kərəmidari ('bricklayers'), and self-ascribed Vlaši ('Wallachians'). The two Čurari samples (Group 4b) share the retention of the person marker 1SG *-om*, the indicative negation in *na*, loan verb integration markers in *-i-*, and *tehe* 'tomorrow', features that are mostly shared with the Vlaši and 'Romanian Roma' (Group 4d), while the three Kərəmidari samples (Group 4c) do not form a more cohesive group in their own right.

Group 5 (Gabor, Šušuwaje and Žuržuvaje) constitutes a remarkably coherent cluster of dialects. It is characterised by the retention of the person marker 1SG *-om* but vowel shift in *čhej* 'girl', consonant de-affrication in *šej* 'girl' and *žukel* 'dog', palatalisation of the initial consonant in *džili* 'song', loss of final *-s* in the remoteness marker *sana* 'you were', in pre-consonantal position *leske* 'for

him' > *le(h)ke*, and in 3SG past tense marker *dikhlja* 's/he saw', but selection of
-s- in intervocalic grammatical morphemes like *lasa* 'with her' and of *s-* in the
copula paradigm, *a*-prothesis in *ašun-* 'to hear', retention of *-l-* in the oblique
definite article, *v*-prothesis in the 3SG pronoun *vov* and in *vušt* 'lip', use of *sode*
'how much', a demonstrative stem with adjectival inflection *kado*, a tendency
to select *-i-* as loan verb marker, and use of *tehe* 'tomorrow'.

Finally, the two samples that carry the self-appellation Rom Mătase or 'silk
makers' (Group 6) are also quite coherent and at the same time distinct from
the other groups. Both show retention of the original vowel as well as the
affricate in *čhaj* 'girl', retention of the affricate in *džukel* 'dog', retention of the
original vowel in the 3SG person ending *-om*, absence of palatalization in *tiro*
'your' but shift of the initial consonant to a fricative in *zili* 'song', loss of final
-s in *phendja* 's/he said', selection of *-h-* in intervocalic morphemes as in *laha*
'with her' and of *h-* in the copula stem, retention of *-l-* in the oblique definite
article, *v*-prothesis in the 3SG pronoun *vov*, and selection of the loan verb inte-
gration marker *-in-*.

7. Geographical diffusion zones

I now turn to a summary discussion of the geographical distribution patterns
of the variants under consideration. The division between the dialects of the
north and those of the south of the country stands out, as does the appearance
of distinct clusters of features in the north-central region and in the northwest.
Dialect boundaries and isogloss clusters reflect boundaries between population
groups that impede social contacts and therefore communication, and thereby
prevent the spread of innovative features of speech from one population to
another. Of what nature were the historical demarcations between Romani
populations in Romania? The partial coherence of samples whose speakers
use the same or similar self-appellations regardless of immediate territorial
coherence suggests that group affiliation plays a role in the diffusion of lin-
guistic structures. We might understand group affiliation in two distinct ways:
The first is a shared origin in a single community that displayed geographical
cohesion within a place or area of settlement, with subsequent dispersion of
some members of the group to other areas where they continued to maintain
their original self-appellation as well as dialect. This interpretation is behind
the notion of 'genetic' dialect branches or groupings. Another interpretation,
which is not necessarily contradictory to the first, is in the sense of a network
of contacts among dispersed communities who share certain interests through
common occupation patterns, values, customs, inter-marriage and so on. The
Kelderash, for instance, might be regarded at both levels: A group originating,
in all likelihood, in the Banat area, which dispersed across the country (and
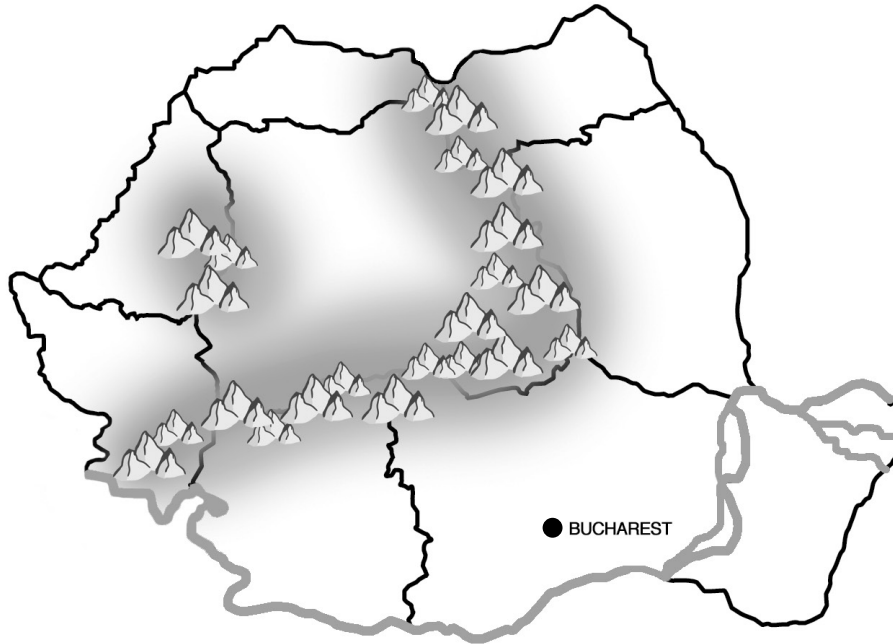
Map 26. Romania: Topography

beyond, to all regions of Europe and the New World), and who continued to maintain tight networks with one another, preserving shared cultural practices including language.

It is tempting to view community boundaries of this kind as the primary source of inter-group demarcation in Romani society, due to the non-territorial nature of the Romani population. But natural and political boundaries may equally play a role in shaping contacts among populations even if they carry out itinerant occupations but remain within well-defined regions. If we consider the general topographical and political landscape of Romania, we can clearly see how the Carpathian mountains form a natural barrier between north and south (Map 26) and the Bihor mountains contribute to the isolation of the Crişana province (Bihor and Arad counties). The historical political division into the three provinces of Moldavia, Transylvania and Wallachia (Map 27) very clearly mirrors the country's physical geography, as does the present-day division into provinces (Map 28). This interplay of geographical and historical political boundaries appears to have been crucial to the mobility of Romani populations, and therefore it had an impact on the nature of contacts between them, promoting or inhibiting the spread of structural innovations in speech.
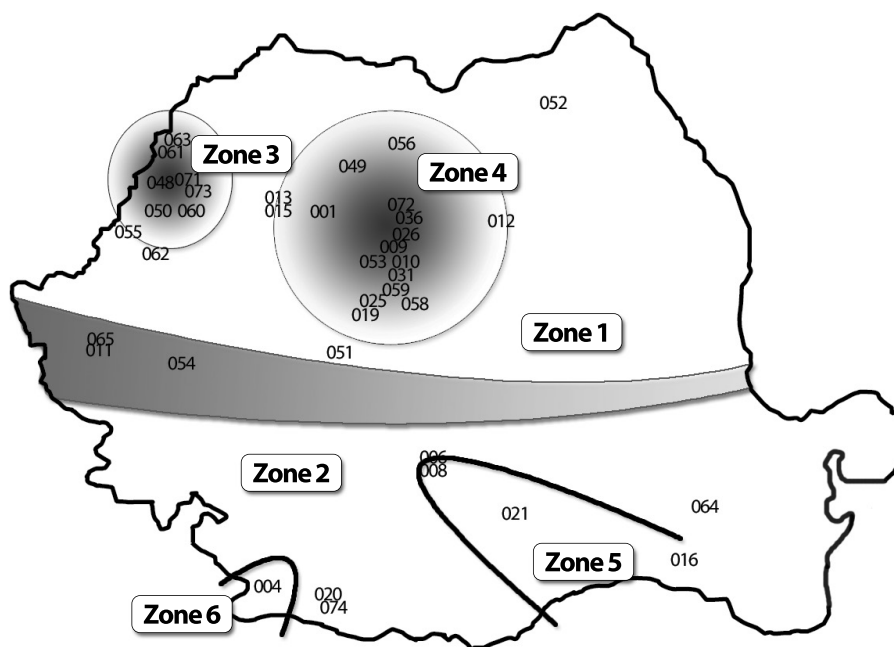
Notwithstanding the proven linguistic connections among groups that share self-appellations, it is thus possible to identify geographical centres of diffusion for linguistic innovations among the Romani dialects of Romania (Map 29).

MAP 27.   Romania: Historical provinces

MAP 28.   Romania: Provinces

Map 29. Difusion zones

These follow the geo-political map rather closely, distinguishing between an area to the north of the Meridional Carpathians (Zone 1), one to the south of the mountains (Zone 2), an area to the northwest of the Bihor Mountains in the Crişana province (Zone 3), and an area to the west of the Oriental Carpathians in Transylvania (Zone 4). Two additional zones appear in the south: The first (Zone 5) comprises the three Spoitori dialects, which are adjoining geographically and may well be regarded as part of a continuum of dialects that stretches southwards into Bulgaria (possibly historical out-migrants from northern Bulgaria). The second (Zone 6) is an isolate consisting of one of the two neighbouring Ursari varieties, which shows considerable distinctiveness.

Our focus is, following the introductory discussion above, on the innovative developments that are centred in and around particular geographical zones. Zones 1 and 2 might be considered 'macro Zones' as they each contain within them smaller zones with more particular developments. I include in the inventories of these two zones only the forms that are present throughout each of the two respective areas. Zone 1 – the area north of the Meridional Carpathians, which separates historical Wallachia from Transylvania (and Moldavia), shows *v*-prothesis in the 3sg pronoun *vov*, the interrogative *sode* 'how much', and *tehe/tese* 'tomorrow'. Zone 1 also shows sporadic and hence apparently incipient tendencies toward palatalization in *čiro* 'your' and de-affrication in *šej* 'girl' as well as frequent selection of the copula stem in *h-*. Zone 2, comprising the

area south of the Carpathians, or historical Wallachia, shows simplification of the historical Early Romani *ndř* (Proto-Romani **ṇḍ*) cluster to *rn* and *r*, the demonstrative *kadava*, and the appearance of a 2PL past tense person marker in *-en* as well as of *ni* as negation marker.

Zones 3 and 4 are both contained within Zone 1 and share those forms that characterise Zone 1. Zone 3 is situated to the north of the Bihor mountains, comprising Bihor and Arad counties along the Hungarian border. This is part of historical Transylvania, a predominantly Hungarian-speaking area, where many of the Romani communities define themselves as 'Hungarian Rom' (Romungro or Ungrika Rom). Here we find the demonstrative *ada/ado*, the past tense 2PL marker in *-e*, selection of *-h-* in intervocalic morphemes (*laha* 'with her'), selection of *-in-* as loan verb marker, and apparently under the influence of neighbouring Zone 4 also sporadic adjectival inflection on demonstratives. Zone 4 is the trans-Carpathian zone, comprising the eastern area of historical Transylvania centred around Mureş county. This area too has a strong Hungarian-speaking population and many Romani communities based here also regard themselves as 'Hungarian Rom'. The area is also home to the Gabor, a Romani group that is said to have consolidated itself over the past century (cf. Jacobs 2012, Olivera 2012). Typical features include the loss of final *-s* in the 2SG present person marker and the loss of *-s-* in pre-consonantal position as in *lehke* 'for him', adjectival inflection on demonstratives, widespread simplification of *-in-* to *-i-* as loan verb marker, and the emergence of *či* as negation marker. Additional features appear sporadically and may be regarded as incipient: the loss of *-s* in the remoteness marker (*sana* 'you were'), the selection of *-h-* in intervocalic morphemes (*laha* 'with her'), and the affrication of the initial consonant in *džili* 'song'.

Zones 5 and 6 are 'micro zones'. Zone 5 consists of the cluster of three Spoitori varieties that are spread across a geographical continuum in the southeast. Typical distinctive features that they share, and which are not shared with other samples from the same area (such as the Kangljari of Ţăndărei, sample RO-064), are the use of *zis* 'day' and the loss of *-l-* in oblique definite article. The cluster also stands out in the area through its retention of the original vowels in *čhaj* 'girl' and the 1SG past tense person marker *-om*, as well as the absence of the past tense 2PL *-en* from two of the three samples, the emergence of a 3SG pronoun *oda* in the same two samples, and the absence, also in two of the samples, of prothetic *a-* in *šun-* 'to hear'. Other features tend to be shared with the southern Zone 2. Zone 6 refers to the particular structural profile of one of the Ursari varieties, that of Maglavit (RO-004). Its most distinctive features are the selection of *-iz-* as loan verb marker (which it shares with the Ursari variety of Sadova) and the demonstrative *adauka*. Other forms are generally shared with the southern Zone 2, though like Zone 5 the dialect stands out through

its retention of the original vowels in *čhaj* 'girl' and the 1SG past tense person marker *-om*.

8. Conclusions

The present study raises a number of methodological issues, and it would seem fair to share those with the reader before proceeding to a discussion of the conclusions. Firstly it is clear that although no previous study has delivered such a comparative dataset of Romani dialect from Romania, the coverage remains very limited. We have no data from the eastern provinces of historical Moldavia, and there are many gaps in the coverage especially of the dialects of southern Romania. For example, we have yet to establish the precise relations between the varieties whose speakers self-identify as Spoitorja and Ursarja respectively and the varieties that surround them in the south. The Romani Dialectological Questionnaire has proven to be a useful tool for the comparative study of Romani dialects, yet in the context of the present article only a small selection of variables could be considered. The full set of existing data from the samples considered in the present study would lend itself to a more exhaustive and detailed analysis were it not for the constraints of space. In relation to both dimensions – the geographical coverage and number of samples, and the range of variables considered in the analysis – the present study must be regarded as a pilot investigation. While the advantages of the questionnaire elicitation method for an effective processing and comparison of data are obvious, it is inevitably limited in its ability to capture variation in speech. Moreover, it is not entirely resistant to misrepresentation through value-triggered production of forms that may be considered prestigious for some reason or other in the interview setting but may be less likely to occur in natural speech. Ideally, then, a dialectological survey would require a more systematic consideration of naturally occurring data.

A comprehensive mapping of the Romani dialects of Romania would also need to take into consideration more detailed speaker meta-data. As far as we have been able to ascertain, all speakers interviewed for the dataset considered here were long-term residents of the locations that they identified as their home communities. But detailed information on their life histories is often missing. We have information on time spent away from the home community from some respondents, but not for all. For most, we also have ethnographic information about customs and occupations, relations with other groups including marriage patterns (i.e. Romani groups that are favoured and disfavoured for marriage partners), religion, and education, as well as the respondents' statement about their group label or endonym, and in some cases also about exonyms (i.e. the way their own group is designated by members of other Romani groups).

All this information was elicited in an interview format and only in some cases was it supplemented by information obtained by the fieldworkers through participant observation in selected communities. Ideally, a detailed dialectological mapping should take into consideration information on networks of contacts, time spent in other locations, the migration history of some families as far as it can be reconstructed, and the values attributed to other groups, in a more intensive and systematic way. There is, in other words, still a need for a comprehensive survey, where historical group affiliation, ethnographic features, the history of nomadism and settlement, cross-regional links among population groups, and structural features of language can be correlated.

Nevertheless, the fact that we find both regional coherences and connections between the variants provided by speakers who gave the same or similar self-appellations in different locations (and may therefore be considered members of the same historical 'group'), speaks for the success and reliability of the method employed here. Had the geographical coverage been seriously constrained by gaps, and had the elicitation of self-appellations produced predominantly arbitrary responses, then one might have expected the distribution of features to be entirely random. This is not the case. Our principal finding is that both group affiliation and spatial distribution shape the linguistic profile of Romani varieties in Romania. This means that historically, structural innovations (changes in the use of linguistic forms) are propagated among speakers and speaker communities both as a result of geographical proximity and through networks of contacts with related groups, and in some cases possibly (though we lack the evidence to prove this in specific cases) through retention of features acquired in one location following a group's migration to another location.

A case in point is the three Spoitorja samples considered here. They occupy, by and large, a coherent area, though at least one variety, in Piteşti (RO-006), is spoken in a community alongside another Romani dialect, that of the Kəldərarja (RO-008). The three Spoitorja dialects share a number of distinctive or even diagnostic features. Some of those are shared with dialects of northeastern Bulgaria, which might point either to an historical migration from Bulgaria into Romania, or simply to a cross-border population continuum through which innovations spread without the displacement of population but merely through continuous contacts and face-to-face interaction. However, not all three Spoitorja dialects are identical in relation to all features discussed in the previous sections. Whereas the samples from Călăraşi and Bolentin Vale have *oda* 'he' and the 2PL past tense person marker -*e*, that from Piteşti, which is situated farthest away from the related dialects of northeastern Bulgaria, has *ov* and -*en*, as do the other Romani dialects of southern Romania. This is a nice illustration of the interplay of group affiliation and spatial location in the shaping of a dialect's structural profile.

Perhaps the typical example of a non-territorial dialect is the three Kelderash/ Kelderari varieties. They are spread right across the country – in Timişoara in the west (RO-065), Suceava in the north (RO-052), and Piteşti in the south (RO-008), yet they share a density of features that generally match those found in the western-centre of the country, which speaks in favour of viewing their present locations as the outcome of migrations from a centre in or around the Banat region. By contrast, the Romungro and Gabor groups each show not just the same endonym but also geographical clustering, by and large, and so group affiliation and territorial coherence go hand in hand.

What we learn from the pilot is that both spatial contiguity and self-ascribed group affinity may shape the structural profile of a dialect. Location may be overridden by group affiliation when the latter implied geographical displacement in recent times and so the retention of a dialect formed elsewhere (a migrant dialect). On the other hand, as in the cases of the Spoitori sample from Piteşti (RO-006), the Ursari sample from Sadova (RO-020), the Ungrika Rom sample from Caransebeş in Timiş (RO-054), as well as others, location tends to override group affiliation as dialects show a tendency to converge with neighbouring varieties. In this way, dialectology can provide a window into the social history of Romani communities, in Romania as elsewhere: neither is spatial location a random factor in the consolidation of a group's identity as manifested by its speech habits and speech fashions, nor do boundaries between contiguous groups appear to constitute obstacles to inter-group communication and the sharing of styles and fashions when it comes to language.

References

Bakker, Peter, and Matras, Yaron. 1997. Introduction. In: Matras, Yaron, Bakker, Peter, and Kyuchukov, Hristo. eds. The typology and dialectology of Romani. Amsterdam: John Benjamins. vii–xxx.

Bakker, Peter. 1999. The Northern branch of Romani: Mixed and non-mixed varieties. In: Halwachs, Dieter W., and Menz, Florian. eds. Die Sprache der Roma: Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internationalen Kontext. Klagenfurt: Drava. 172–209.

Bochmann, Klaus. 1999. Der Einfluss des Romani in Südosteuropa am Beispiel des Rumänischen. In: Hinrichs, Uwe. ed. Handbuch der Südosteuropa-Linguistik. Wiesbaden: Harrassowitz. 701–8.

Boretzky, Norbert. 1999a. Die Gliederung der Zentralen Dialekte und die Beziehungen zwischen Südlichen Zentralen Dialekten (Romungro) und Südbalkanischen Romani-Dialekten. In: Halwachs, Dieter W., and Menz, Florian. eds. Die Sprache der Roma: Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internationalen Kontext. Klagenfurt: Drava. 210–76.

Boretzky, Norbert. 1999b. Die Verwandtschaftsbeziehungen zwischen den Südbalka-nischen Romani-Dialekten. Mit einem Kartenanhang. Frankfurt am Main: Peter Lang.

——2000. South Balkan II as a Romani dialect branch: Bugurdži, Drindari, and Kalajdži. Romani Studies, Series 5, 10: 105–83.

——2003. Die Vlach-Dialekte des Romani. Wiesbaden: Harrassowitz.

——2007. The differentiation of the Romani dialects. Language Typology and Univer-sals 60: 314–35.

——and Igla, Birgit. 1993. Lautwandel und Natürlichkeit. Kontaktbedingter und endogener Wandel im Romani. (Arbeitspapiere des Projekts "Prinzipien des Sprachwandels" 15). Essen: Universität GH Essen, Fachbereich Sprach- und Lite-raturwissenschaften.

—— ——2004. Atlas der Romani-Dialekte. Teil 1: Vergleich der Dialekte; Teil 2: Dialektkarten mit einer CD Rom. Wiesbaden: Harrassowitz.

Constantinescu, Barbu. 1878. Probe de limba sçi literatura Tçiganilor din România. Bucuresçti: Typografia SocietaΣtçii Academice Române.

Courthiade, Marcel. 1998. The dialect structure of the Romani language. Interface 31: 9–14.

Drimba, V. 1992. Imprumuturi românesti din limba tiganeasca. Studii şi Cercetări Lingvistice 43: 173–90, 253–78.

Elšík, Viktor and Matras, Yaron. 2006. Markedness and language change: The Romani sample. Berlin: Mouton de Gruyter.

Gaster, M. 1931. Rumanian Gypsy folk-tales. Journal of the Gypsy Lore Society, Third Series, 10(4): 153–71; 12(4): 166–89; 15(1): 10–20; 15(4): 160–81; 17(3): 58–66.

Gilliat-Smith, Bernard J. 1915. A report on the Gypsy tribes of North East Bulgaria. Journal of the Gypsy Lore Society, New Series, 9: 1–54, 65–109.

Graur, Alexandru. 1934. Les mots tsiganes en roumain. Bulletin Linguistique Romane 2: 108–200.

Jacobs, Fabian. 212. Bewegliche Kulturen. Eine kongruente Geschichte der Gabor in Siebenbürgen. Leipzig: Eudora Verlag.

Juilland, Alphonse. 1952. Le vocabulaire argotique roumain d'origins tsigane. Cahiers Sextil Puşcariu 1: 151–81.

Kaufman, Terrence. 1979. Review of Weer Rajendra Rishi, Multilingual Romani Dictionary. International Journal of the Sociology of Language 19: 131–44.

Labov, William. 1994. Principles of language change: Internal factors. Oxford: Blackwell.

Leschber, Corinna. 1995. Romani lexical items in colloquial Rumanian. In: Matras, Yaron. ed. Romani in contact: The history and sociology of a language. Amster-dam: John Benjamins. 151–76.

Lucassen, L. 1996. Zigeuner: Die Geschichte eines polizeilichen Ordnungsbegriffes in Deutschland (1700–1945). Cologne/Vienna: Böhlau-Verlag.

Marushiakova, Elena and Popov, Vesselin. 2013. 'Gypsy' groups in Eastern Europe: Ethnonyms vs. professionyms. Romani Studies, Series 5, 23: 61–81.

Matras, Yaron. 1999. The speech of the Polska Roma: Some highlighted features and their implications for Romani dialectology. Journal of the Gypsy Lore Society, Fifth Series, 9: 1–28.

——2002. Romani: A linguistic introduction. Cambridge: Cambridge University Press.

——2005. The classification of Romani dialects: A geographic-historical perspective. In: Halwachs, Dieter W., Schrammel, Barbara, and Ambrosch, Gerd. eds. General and applied Romani linguistics. Munich: Lincom Europa. 7–26.

——2010. Romani in Britain: The afterlife of a language. Edinburgh: Edinburgh University Press.

Miklosich, Franz. 1872–1880. Über die Mundarten und Wanderungen der Zigeuner Europas X–XII. Vienna: Karl Gerold's Sohn.

Okely. Judith. 1983. The Traveller Gypsies. Cambridge: Cambridge University Press.

Olivera, Martin. 2012. The Gypsies as indigenous group: The Gabori Roma case in Romania. Romani Studies, Series 5, 22: 19–33.

Pott, August. 1844–1845. Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache. Halle: Heynemann.

Sampson, John. 1926. The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood (repr. 1968). Oxford: Clarendon Press.

Turner, Ralph L. 1926. The position of Romani in Indo-Aryan. Journal of the Gypsy Lore Society, Third series, 5: 145–89.

Urech, Evelyne and van den Heuvel, Wilco. 2011. A sociolinguistic perspective on Roma group names in Transylvania. Romani Studies, Series 5, 21: 145–60.

Ventcel', Tatjana V., and Čerenkov, Lev N. 1976. Dialekty cyganskogo jazyka. Jazyki Azii i Afriki I. Moskva: Nauka. 283–332.

Willems, Wim. 1997. In search of the true Gypsy: From enlightenment to final solution. London: Cass.