RomIdent Working Papers

Paper No.15

Romani dialect classification revisited

Yaron Matras (University of Manchester)

2010

Appeared in print as: Chapter 2, Matras, Y. 2010. Romani in Britain: The afterlife of a language. Edinburgh: Edinburgh University Press

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

The project RomIdent is financially supported by the HERA Joint Research Programme (www.heranet.info) which is co-funded by AHRC, AKA, DASTI, ETF, FNR, FWF, HAZU, IRCHSS, MHEST, NWO, RANNIS, RCN, VR and The European Community FP7 2007-2013, under the Socio-economic Sciences and Humanities programme.

EUROPEAN COMMISSION
European Research Area

SEVENTH FRAMEWORK PROGRAMME

Funded under Socio-economic Sciences & Humanities

Yaron Matras

Romani dialect classification revisited

TOWARD A CHRONOLOGY OF ROMANI MIGRATIONS

No historical record is known to date that can be linked unequivocally with the ancestors of today's Romani population, nor is there any specific mention of the Rom or their language, or any other specific citation from the Romani language that would allow us to link any general historical events with the specific population known as Rom. Nonetheless, speculation about the participation of an ancient Romani population in various state institutions, migrations, battles, and so forth have been thriving since the second half of the nineteenth century (and partly even earlier). It has recently been given a renewed impetus through the work of a circle of Romani political activists and their supporters engaged in trying to disseminate an historical narrative that is intended to portray the Rom in a way that challenges popular images about them. The principal argumentation line in these essays is that the ancestors of the Roms had occupied a high and privileged status in their country of origin, but were deprived of this status under various circumstances and forced into exile in Europe, where marginalisation and stigmatisation first emerged (see e.g. Hancock 2002, Marsh 2008). Critics of this thesis are often branded justifiers of the social exclusion that is imposed upon the Rom (see Hancock 2008).

While the reasons for emigration away from India remain inaccessible to investigators, there is some circumstantial evidence to be considered. It concerns primarily the presence of other groups with Indic languages and service economies outside of India, such as the Jat of Afghanistan, the Parya of Central Asia, the Lom of the Caucasus, and the Dom of the Middle East. These groups used to, and in most cases continue to specialise in the

same range of traditional occupations as the Rom of Europe, i.e. in a mobile service economy. The groups maintain a tight-knit community structure based on kin loyalty even when settling in urban districts. The names they give to themselves derive from Indian caste names and many correspond to existing caste names still found in India today, such as that of the *ḍom*, who specialise in a mobile service-economy. The names they use to denote the outside, settled population, are often cognate with the Romani word *gadžo* and point to a shared conceptual understanding of the roles of outsiders (though not necessarily to a shared ancient language, beyond the fact that all groups speak related modern Indo-Aryan languages). The presence of these various groups outside of India confirms an overall phenomenon of emigration from India of specific caste groups and the maintenance of caste-like identity even after the breakaway from the actual caste-based social system of the Indian subcontinent.

Like the Rom and the Middle Eastern Dom, some of these populations, e.g. the Parya of Afghanistan and Tadjikistan or the *ḍum* or the Hunza Valley in northern Pakistan, speak Central Indian languages. In connection with the linguistic changes discussed above and the periods in which they are documented in writing, the prevailing assumption continues to be that the Rom are descendants of a population belonging to mobile, service-providing castes who migrated from Central India northwards in early medieval times, sometime around the middle of the first millennium CE. They remained in the Indian Subcontinent long enough for their language to manifest the typical features of the major transitions to Early New Indo-Aryan, thus possibly until the ninth or tenth century CE, at which point the population migrated once again, settling eventually in the eastern regions of the Byzantine Empire, in or around present-day eastern Anatolia bordering the Caucasus.

An origin in service-providing castes in India could well be reconcilable with the view that the Rom left India as camp followers who made their living by providing crafts and services to military forces. This idea was first put forward by De Goeje (1903), and has since offered a middle-of-the road explanation for the migrations of the Rom: They were neither aimless wanderers, as they are often portrayed in the non-specialised literature, nor a prestigious caste of warriors and priests, as portrayed by some Romani activists (e.g. Kochanowski 1994). Some of the early loan vocabulary, such as the term *koraxaj* (also *xoraxaj* used for 'Turks' and in some varieties of Romani generally for 'foreigner', from the name of the Turkic *Karakhanide* empire in Central Asia, the Greek word *kris* ('judgement') used to denote traditional Romani courts, and Iranian *baxt* 'luck, good fortune', hint at a world view that was informed and inspired by the proximity to the military might of the Muslim Turkish armies, by the spirituality of neighbouring peoples, and by the social organisation of Byzantine Greek society. Quite possibly, the linguistic evidence of considerable contact with the populations of eastern Anatolia and the Caucasus and of intense contact with late medieval Greek-speaking society reflects a period during which the dependency on a civilian population was restored and the Rom took on a more permanent economic role in the eastern part of the Byzantine Empire. In any event, the intensity of the Greek linguistic influence prompts speculation that bilingualism and a stable position in Greek-speaking society may have lasted for many generations, perhaps as long as two or three centuries.

Early attestations of Gypsies that can be interpreted as references to the Rom rather than to any other itinerant population appear in the Balkans in the late fourteenth century and testify to the immigration by that point in time of a Romani population westward. Some authors have attributed this movement to the gradual disintegration of the Byzantine Empire

and the threat of Turkish invasions. The fate of the Romani population of eastern Anatolia remains uncertain. The only known Romani-speaking populations in the region are the descendants of relatively recent immigrants from the Balkans, who arrived after the establishment of the Turkish Republic in the 1920s. But there are at least three other itinerant populations in the area: The Domari-speaking Dom, the Armenian-speaking Lom (who retain an Indic-derived vocabulary), and the indigenous Abdal, who have an in-group lexicon based partly on Domari as well as Romani. It is possible that the Rom moved out of the area, and that any remaining groups assimilated into neighbouring peripatetic communities or into the settled population. The Romani-speaking Zargar of Iranian Azerbaidjan are believed to have migrated eastwards to their present location from northern Greece sometime in the eighteenth century, indicating that there was at least some Romani presence in eastern Anatolia since medieval times, albeit of a transitory nature.

Much like the immigration of Roma into the eastern Byzantine Empire, the reasons for their migrations in the fourteenth century into western and northern Europe remain unclear. The only hypothesis put forward so far suggests an attempt to escape from the turbulence surrounding the collapse of the Byzantine Empire and the gradual rise of the Ottoman state in its place. While this might have certainly triggered a movement out of the area by some, we must also note that the majority of the Romani population remained in the Balkans under Ottoman rule.

In all likelihood, judging by the chronicles that depict their arrival in numerous European towns, migration took place in households comprising groups of several dozen persons of all ages. The fact that during the early fifteenth century similar narratives were cited from Rom in different places regarding the reasons of their travel – a pilgrimage from Egypt being one of the more frequent stories recorded – and that possibly even safe conduct

letters were copied or otherwise shared between groups arriving in different locations at different times, indicates that there were contacts between individual households and perhaps even stable social networks across regions and locations. Nonetheless, we must assume that during this period it was the family households, and their closest kin relations with whom they kept contact, that constituted the smallest social unit. Travelling in social isolation from the settled population, these units also constituted tiny speech communities. In the absence of any reports on large-scale settlements or the sudden arrival of larger Gypsy communities in any one location or region, we must assume that the period of migrations lasting well into the late fifteenth century was characterised by the random arrival and settlement of individual households in and around western and northern Europe. Members of kin-related households may have followed and joined those early migrants in places that proved favourable as far as the reception on the part of the settled population and the prospects of earning a livelihood in the short term were concerned, thus creating larger communities.

THE PERIOD OF DIALECT FORMATION

Settlement at a particular location usually meant that Rom were tolerated and given the opportunity to engage in trades for which there was local demand, without competing with local tradesmen. In the Ottoman and Habsburg Empires as well as in the Romanian principalities a general division emerged between itinerant Rom, whose trades required mobility between client populations, and settled Rom, who provided a range of services to local land owners as well as to villagers, including seasonal field labour. The ties and dependencies that were formed between these latter Romani communities and the settled population were of course particularly strong, even if the social distance between them remained enormous. To the west and north of the Habsburg monarchy, however, there is little

evidence of any large-scale integration of a Romani work-force into local economies. The Romani populations of these regions remained, often until the twentieth century and in Britain until this very day, a mobile service economy. When we speak of 'settlement' in this latter context we therefore mean the adoption of a routine pattern of trade and services and a network of clients and opportunities based in a particular region and embedded into the economic, social, geographical and linguistic context of that region.

The period beginning in the middle of the sixteenth century and lasting well into the enlightenment and the emergence of modern nation-states is well known as a period of anti-Gypsy persecution, as testified by hundreds of royal edicts, regional decrees and police memorandums from across the European continent. But the absence of security and stability in the lives of the Romani populations during this period does not necessarily contradict a gradual process of accommodation to the particular social environment of their territory of settlement. Indeed, in a period during which long-distance travel will have been extremely difficult and even dangerous due to severe limitations, local and regional networking will have been essential for survival. This includes both networking among the Romani households and clans within a region or territory, and the cultivation of trade contacts with the local settled population.

Despite the hostility of the state and probably a large portion of the population too, the descendants of Romani immigrants in individual regions underwent a kind of integration process: They acquired the regional languages and, without allowing them to interfere with their own group-internal set of beliefs and values, they adopted each region's religion and even some of the regional religious practices, such as pilgrimage to recognised locations and participation in certain festivities. They also maintained regional networks through fairs and regular regional travel, thereby intensifying their interaction with Rom from the same and the

immediately neighbouring regions. At the same time contacts with Rom in more remote locations were gradually lost. Romani communities thus began to develop their individual local identities without necessarily abandoning inherited traditions. Apart from the differences acquired through partial accommodation to the external environment, differences emerged also in internal organisation forms: dress traditions, conflict-resolution institutions and forms of leadership, customs surrounding marriage and death, and the precise forms of implementation of an inherited code of honour and spiritual morality that tightly regulates the mode of interaction among members of the community.

Documentation of the Romani language is rather sporadic until the early eighteenth century, but becomes more prolific during this period with growing interest in the movements, social and family networks, customs and ultimately also in the origins of Gypsies on the part of law enforcement agencies as well as scholars and academics. Already the earliest specimens of Romani, such as those by Borde (1542), van Ewsum (ca. 1560), Vulcanius (1597), Evliya (1668) and Ludolf (1691), when compared with one another give a picture of dialect differentiation that greatly resembles the one that is familiar to us from contemporary observations. This picture becomes even more elaborate, covering many new regions, thanks to numerous compilations circulated in the first half of the eighteenth century. It confirms in yet more detail that by this period the major structural developments responsible for dialect differentiation within Romani had already taken place.

One possible way to interpret this is to assume that distinct dialects had already formed prior to migration and settlement in the individual regions. As indicated above, no language is entirely uniform and variation will have occurred within Early Romani too. It is likely that the speech forms of the individual clans that settled in various regions were not identical to one another in each and every structural aspect. However, since the westwards

migrations were essentially migrations of extended kinship groups who set out to seek favourable opportunities to engage in local or regional service trade, and not one of tribes who took over entire territories nor a coordinated re-settlement of populations from one particular district into another, it would require an extraordinary coincidence for larger territories outside the southern Balkans to be targeted exclusively by Romani families whose speech forms resembled one another, while families with different speech forms were attracted to other territories. The hypothesis of an Anatolian genesis of dialect differentiation in Romani as put forward by Boretzly & Igla (2004) and Boretzky (2007) is therefore difficult to accept. For one, it is impossible to corroborate such an hypothesis through evidence of any ancient dialect differentiation within Romani found in Anatolia itself. But it is equally difficult to imagine how linguistically coherent sub-groups, even if they had existed in Anatolia, might have coordinated their migrations and settlement in western and northern Europe in such a way as to ensure the dominance of a particular Romani dialect in a particular area of settlement in the west. Moreover, many of the structural differences among the present-day dialects of Romani owe their existence directly or indirectly to the influences of the respective co-territorial languages. This concerns not just obvious loanwords, but also changes in the sound system and word stress patterns, lexical semantic developments, and changes in the productivity and frequency of particular inherited morpho-syntactic patterns.

For example, the truncation of initial syllables, common to the Romani dialects of the west, affects the internal or inherited Romani structural component: consider German Romani *glan* 'in front' from **anglal*, *vela* 'he/she comes' from **avela*, and *pre* 'above' from **opral*. There is little doubt that this development is set in motion through the adoption of the Germanic initial word stress, and that such a development necessarily followed several generations of Romani-German bilingualism. But even other, genuinely internal changes

might have been propagated more easily in the isolated, tight-knit, household-based Romani communities once these began to develop local identities of their own, along with shifting centres of prestige and targets of imitation. Finally, the present-day linguistic landscape of Romani testifies to the successive spread of innovations from a variety of different centres of diffusion, in different directions and to different extents, thus forming a complex web of intersecting territorial isoglosses. Such patterns cannot possibly emerge as a result of the import of coherent dialects into clearly demarcated zones. We must therefore conclude that the most prominent differentiation features separating present-day Romani dialects emerged after settlement and the adoption of a local Romani group-identity, in other words, from the early sixteenth century onwards, and that they were well in place by the time documentation of Romani proliferated in the early eighteenth century.

Variation within Early Romani

While no language is entirely uniform, we lack any concrete evidence about any major dialect differences in Early Romani prior to the dispersion and settlement in Europe. There is, however, evidence that some processes that led to dialect differentiation were set in motion at a period prior to the dispersal and settlement in present-day locations. The evidence comes from the historical reconstruction of a variable coupled with an evaluation of the present-day geographical distribution of the complete set of forms representing that variable. The absence of clear geographical patterning, or a random distribution of the forms, in conjunction with the absence of any particular trigger for the various local developments, will indicate that variation existed prior to the dispersion and the formation of the dialects.

A good example is the adoption of the so-called prothetic segments *v-* and *j-* in word-initial position preceding *a-*. While a similar development in positions preceding *u-/o-* and

i-/e- respectively can be regarded as phonologically conditioned (thus *ušt* 'lip' > *vušt*, *iv* 'snow' > *jiv*), Turner (1932) had identified the roots of the development preceding roots in *a-* in a morphological process, namely the fusion of the demonstrative-turned-definite article m. **ov* f. **oj* with the following noun. As evidence, Turner cites the initiation of the process in those three words in which consonant prothesis is uniform across all dialects of Romani: m. *v-ast* 'hand' < MIA *(h)ast*, f. *j-ag* 'fire' > MIA *agi*, and f. *j-akh* 'eye' < MIA *akhi*. It is clear that such a process could only have been set in motion after the emergence of definite articles, and so after contact with Byzantine Greek and therefore in the two centuries or so prior to the dispersal of Romani populations through Europe. At the same time, the initiation of the process will have begun before the definite article form was reduced to its present-day forms m. *o*, f. *i/e*, which is likely to have been long before the European immigration since no present-day dialects retain full consonantal forms for the complete definite article paradigm. We are thus dealing in all likelihood with a development that began in Early Romani. During the common phase it was firmly adopted in the above three nouns, attaching variably to an additional small number of masculine nouns. Three of those are *v-ařo* 'flour', *v-angar* 'coal' and *v-andřo* 'egg'. The present-day distribution of the forms tends to follow a centre-periphery pattern, with various geographical peripheries (in changing constellations, depending on the individual word in question) selecting the more innovative form in *v-* while the centre ends up rejecting the innovation and opting to generalise the more conservative form. We thus end up with an interaction between the forces of geographical diffusion and the inheritance of variation.

Another case of such interface is the generalisation of copula stems containing the extension *-in-* (*s-in-om* 'I am' etc.) in the southern European periphery, comprising the Romani dialects of western Bulgaria, Macedonia, and on both sides of the Adriatic coast,

while the dialects of Greece show a mixture that can be taken to represent the original variation in Early Romani (see Matras 2004: 102). Apart from the presence of both types of copula form, with and without *-in-*, in the present-day Romani dialects of Greece, additional evidence for historical variation comes from the occasional appearance of copula forms in *-in-* in isolated paradigm positions in other, more remote dialects, such as Finnish Romani and Eastern Slovak Romani, where they appear in the 3rd person present (*hin*, *hine*). The origin of the formation is likely to be in the re-interpretation of the past-tense stem of the mono-consonantal root *s-/h-* as a present-tense form. The augment in *-in-*, originally an adjectival-participial ending, belongs to the pool of perfective endings that are favoured with ambiguous past-tense forms, such as those that are based on plain participles (3rd person forms, in particular third person plurals), passives and inchoatives, verbs expressing emotion, and a small set of mono-consonantal verb stems including *s-/h-* 'to be', *d-* 'to give' and *l-* 'to take'. Variation among the dialects thus reflects a stage of variation within Early Romani, where the choice of an augment was optional. After dispersion and settlement, individual varieties of the language opted for a stable setup. The generalisation of forms in *-in-* across a southern belt reflects the region-specific diffusion of a solution to an inherited option.

The co-existence of two separate copula stems in Romani – in *s-* and in *h-* – is itself a further illustration of the way Early Romani variation is inherited into the dialects. There are basically three continuation options. A group of dialects in Macedonia and Kosovo show both sets, directly continuing the inherited variation. Other dialects opt for either one consonantal root or another. In most regions we find that *s-* prevails, but the *h-*set is generalised in the Romani dialects of Germany and neighbouring regions and is also attested in individual dialects in Transylvania and in northern Greece. The third option is to adopt a mixed paradigm, where forms in *h-* appear in individual slots, most likely in the 3rd person

present, and sometimes exclusively in enclitic position. Such mixed paradigms are attested in dialects as far apart as Montenegro, Slovakia and Finland, showing that there are instances of inherited variation for which no geographically coherent preference is visible. Instead, variation appears to be conditioned by local factors, and the geographical distribution is to some extent random. This pertains especially to the choice of lexical items. While some lexical isoglosses split the entire Romani-speaking landscape into large coherent zones, for numerous words neighbouring dialects have conflicting preferences. Most prevalent is a dense variation of preferences in southeastern Europe and the Balkans, the historical diffusion centre of all Romani dialects. Here we find, for instance, forms like *mami*, *baba* and *phuridaj* for 'grandmother', or *men* alongside *kor* for 'neck', side by side in the same region.

The Balkans are also home to numerous different realisations of the historical retroflex cluster *ɳɖ*, including the preservation of a retroflex sound (*maɽo* 'bread'), of various options of a non-retroflex cluster (*mandro*, *marno*, *mando*, *manglo*, etc.), and of the simplex *r* that is otherwise prevalent throughout the north of Europe (*maro*). The density of different forms in close proximity to one another makes it quite easy to imagine the prolonged co-existence of different variants continuing the old cluster *ɳɖ* already before the migration westwards. Finally, Early Romani appears to have shown palatalisation of dental and velar stops in positions preceding /i/, the results of which survive often in diverse, word-specific realisations of the original segments. Thus *(o)gi* 'soul, heart' may continue as *gi*, *dži*, *zi* and so on. Preference toward one or the other continuation of a palatalised segment in one word does not necessarily imply a preference for a similar solution in another word. The outcome is a proliferation of combinations that are often specific to a particular local or regional speech community.

LOCAL AND REGIONAL CHANGES

While we are able to postulate Early Romani variation for some cases of cross-dialectal differences found today, other cases appear to be the outcome of simplification and levelling processes acting upon the full and coherent Early Romani inheritance, albeit in different ways in different communities. Simplification and functional decline characterise the fate of various grammatical devices, among them the use of enclitic nominative pronouns of the set *lo, li, le* , the use of gerunds in *-indo(j)*, the use of Greek-derived 3SG concord ending *-i* and of Greek-derived numerals from 'thirty' onwards (which are often replaced either by internal formations or by subsequent borrowings), and the collapse of Greek-derived nominal inflection endings and verb integration affixes.

Many changes in the dialects are, of course, directly induced by language contact, and take on different shapes in accordance with the source language. Typical functional domains in which word-form or morpheme borrowings occur are comparative and superlative markers in adjectives, nominative plural endings on the noun, indefinite markers and indefinite word forms, conjunctions and discourse markers, conditional and interrogative particles, modal verbs indicating necessity and ability, and prepositions such as 'against', 'between', and 'without'. Romani dialects in contact with Russian, Polish, Czech and Slovak, tend to borrow the full set of so-called aspectual (aktionsart) prefixes. Other areas of morpho-syntax are frequently subject to restructuring as a result of contact. They include the productivity of definite and indefinite articles, the semantic distribution of nominal cases, the generalisation of a single form of the verb in modal complements (infinitive), changes in word order (affecting especially the position of object pronouns), the productivity of verb derivational morphology, and the lexical-semantic expression of aktionsart. Typical contact-induced

changes in phonology include the acquisition of vowel length, changes in stress patterns, the acquisition of additional phonemes and consonant palatalisation.

All this adds up to an enormous pool of potential innovations and so to numerous possible outcome scenarios of a local and regional character. As in any other language, every linguistic structure is potentially open to change and innovation in any community of speakers who use Romani. Many of these changes will remain confined to the domain of lexical preference in the context of family communication, and so they will have little affect on the speech of entire communities. Others will be strongly shaped by the contemporary contact language and will therefore spread more or less along pre-determined lines defined by the nature of the multilingual setting. In between, changes might emerge locally and receive acceptance within a limited range of social interaction networks, encompassing perhaps a group of settlements or even a group of related families who interact with one another across greater distances. While all these innovations will contribute to shaping the speech variety of each and every individual and community, they are of little use toward an overall classification of Romani dialects due to the rather limited distribution that they receive. Any approach that chooses to focus on each and every local innovation without identifying a hierarchy of more and less prominent features for comparison will inevitably end up having to define the idiolects of individual speakers as potentially independent varieties. Having identified some of the areas that are particularly prone to variation in Romani, I shall therefore proceed in the next section to outline some of the more prominent developments that receive wide-scale diffusion across larger geographical spaces, and which slice through the Romani speaking landscape and divide it into larger zones, i.e. into units that provide meaningful indications of historical networks of contacts among speaker communities during the relevant periods.

TERRITORIAL DEVELOPMENTS AND MAJOR ISOGLOSSES

There is ongoing discussion in Romani linguistics whether to regard differences among dialects as territorial, i.e. conditioned by the location of a dialect relative to the geographical spread of a particular structural innovation, or as 'genetic'. The 'genetic' metaphor suggests that certain features must be taken for granted due to ancestry rather than be understood as the outcome of a gradual process of acquisition involving exposure, accommodation and finally adoption of the feature in question. Such an impression of the *Vlax* Romani dialects of northeastern Bulgaria had led Gilliat-Smith (1915) to classify them as 'genetically' distinct from other co-territorial varieties and to postulate that they were not formed in their present location, but had been brought into the region as a result of an immigration of Rom from Romania (specifically Wallachia). The noticeable presence of Vlax dialects in urban centres all across Europe, the outcome of later migrations from Transylvania and Banat from the mid-nineteenth century onwards, made the distinction between Vlax and non-Vlax dialects of Romani a pertinent one in subsequent work on Romani dialectology.

The dispersion of the Vlax, coupled no doubt with the very fact that Romani itself is known to have non-European 'genetic' origins, has created somewhat of a fixation within the study of Romani on interpreting distinctive structural features as proof of a primordial displacement rather than as the outcome of a process of acquisition through interaction and exchange (as an example see Boretzky 2007). In this section I will briefly show how the present-day distribution of major structural features within Romani in geographical space must be interpreted as the outcome of a series of major changes that spread across chains of neighbouring communities, each change sub-dividing the entire Romani-speaking landscape into a limited number of zones. While the spread of some developments follows common pathways creating clusters of linguistic boundaries or isoglosses, the patterns that emerge are

by no means uniform. Instead, isoglosses intersect in numerous different ways in a complex matrix. This matrix can be read as an illustration of the ever-evolving targets of social contacts, prestige and imitation that lead speakers from one community to adopt selected features of speech that arise in a neighbouring community.

The geographical diffusion model goes hand in hand with an appreciation of historical migrations of population groups. We must reconstruct the original geographical context for those groups that are known to have migrated to their present locations after the formation period of the dialects had ended, that is from the eighteenth century onwards. To be sure, subsequent changes will have taken place in all dialects, but these must be examined separately. The so-called Southern Vlax dialects that spread among local, settled dialects of the southern Balkans must therefore be examined together with the closely related varieties of the adjoining regions to the north, namely Serbia and Banat and the continuum that they form into the Vojvodina region in the west and Wallachia and Transylvania to the east and north. Northern Vlax dialects that left the Transylvania and Banat regions in the nineteenth century must similarly be considered migrant dialects.

Second, there are indeed instances where shared structural features may confirm a breakaway of one group from another and its migration to a remote location. There is little doubt that the similarities between the speech forms of the Lithuanian Rom and those of the Russian Rom of the Urals will have emerged prior to the arrival of Rom in the Urals. They were not, in other words, a result of gradual changes to which a Romani population in the Ural and been exposed and which it adopted, but the result of an exchange that took place while the two groups had been in much closer proximity to one another, somewhere closer to the Russian Baltic coast, and were later on brought to the Urals by a population of migrants. Similarly, features shared by the Romani dialects of Germany and those of Finland are less

likely to have diffused gradually from their emergence centre in Germany to reach a Romani population that had already been settled in Finland. It is much more plausible to attribute those features to a period during which the ancestor population of the Finnish Roma resided in or close to Germany. They were then carried in the speech of this group when it migrated to its present location.

Keeping our eyes open for such issues, the plotting of dialectal features on the map allows us to make the following generalisations about the geographical diffusion patterns of structural innovations among the dialects of Romani.[1] A major division is visible between the dialects of western and northern Europe, and those of southeastern Europe. The dividing line (also referred to as the 'Great Divide'; see Matras 2005) runs roughly between northern Ukraine in the east and the northern tip of the Adriatic coast in the west. It is a cluster of isoglosses, not a single line, and the precise path of individual isoglosses varies somewhat. Some divisions run far enough to the north as to include southern Poland and the whole of Slovakia as well as eastern Austria and Slovenia on the southern side of the line; others have a course that cuts across the region farther to the south, leaving either just northern Slovakia or sometimes the whole of Slovakia on the northern side of the line. Transition zones are not uncommon in this area even when we examine just a crude sample that does not take into consideration the full density of settlements or communities. Studies focusing on particular sub-regions are likely to find even greater variation on both sides of major isoglosses.

[1] The discussion presented here is based on an ongoing evaluation of data from over 200 locations in Europe, compiled and stored as part of the Romani Morpho-Syntax Database (RMS). The resource is freely accessible online: http://romani.humanities.manchester.ac.uk/. For an additional comprehensive source of dialect maps for Romani see Boretzky & Igla (2004).

Even at a superficial glance it is quite clear that the Great Divide reflects the political division and conflict zone between the Habsburg Monarchy to the north and the Ottoman Empire to the south, during the crucial period of dialect formation that followed Romani settlement, in the sixteenth and seventeenth centuries. The border separated two major Romani population centres and made it impossible for structural innovations that emerged on one side to be carried over and diffused on the other side. Here too, the presence of a geographical demarcation line does not exclude the possibility that population movements were partly responsible for shaping the precise distribution picture. The absence of a direct continuum between the Arli-type (Southern Balkan) dialects of Macedonia, Kosovo, and Albania, and the (Southern Central) dialects of Slovenia, eastern Austria and Hungary, which share some similarities with the first group, might be interpreted as reflecting a displacement of Rom from present-day Croatia northwards, brought about quite possibly as part of the evacuation of civilian populations loyal to the Habsburgs during the seventeenth century. Rom from other regions further to the east will have later moved into the region, carrying with them the Vlax-type dialects that are spoken in Croatia today. Such a scenario must still be confirmed with the help of historical documentation. But even if confirmed it would not question the validity of the geographical diffusion model, but would merely add circumstances that could help explain why the division is so clearly pronounced over such a relatively dense zone, and why some of the isoglosses run much farther to the north than the old political borders and conflict zones, thus dividing southern and northern Slovakia.

The Great Divide or North-South division between Romani dialects represents the spread of a series of unconnected structural developments. Germany appears to be the epicentre for a series of innovations on the northern side. Syllable truncation is one of the typical developments in this region, triggered in all likelihood by a shift to word-initial stress

as a result of Romani-German bilingualism. The north thus has *mal* 'friend' for *amal*, *khar-* 'to call' for *akhar-*, *sa-* 'to laugh' for *asa-*, often *kana* 'now' for *akana*, and more. Further developments include a preference for prothetic jotation in selected words, among them *jaro* 'egg' and the 3rd person pronouns *jov* 'he', *joj* 'she', *jon* 'they', and the simplification of the historical cluster *ṇḍ* to *r* in words like *maro* 'bread', *miro* 'my' and *jaro* 'egg', while the south maintains a proliferation of cluster combinations that continue the historical sound (see above). The remarkable coherence of the entire northern area, from Britain to Finland, the Baltics and northern Russia, in relation to these features might be interpreted as an earlier spread among the dialects at a time when their areas of settlement where still closer to one another and social networks among them were tighter, or indeed prior to the split of an earlier group settled around the German-Polish contact area into several sub-groups which then migrated in different directions. Note that the Romani dialects of the Iberian peninsula tend to remain conservative with respect of some of these features, indicating that they were not part of the network of contacts that enabled their diffusion in the north.

A number of typically northern developments fail to reach the extreme northern periphery of Finland and appear to have been adopted after the breakaway of the Scandinavian sub-group. They include the loss of the preposition *katar* 'from', which is retained in both British and Finnish Romani, and the assimilation of intransitive verbs of motion and change of state into the dominant verb inflection and the disappearance of gender-inflected past-tense forms or active participles of the type *gelo* 'he went' *geli* 'she went' (retained in Finnish Romani).

A series of lexical preferences spread throughout the north, based on inherited variation that often continues in the south. Thus the north has *xač-* 'to burn' (in the south *phabar-*) and *stariben* 'prison' (*phanglipe* in the south, but also in Finnish Romani), as well

as *angušt* ‘finger’ (*naj* in the south), derivations of *gi* for ‘heart’ (*ilo* in the south), and *men* ‘neck’ (*kor* in the south). The south, in turn, has its own non-conforming periphery usually comprising an area along the Black Sea coast and in Greece, and it is here that both *angušt* and *gi* are also preferred, while *men* is found sporadically in the Balkans alongside *kor*.

In the south, the epicentre of innovation appears to be Romania and adjoining regions in all directions. Prominent southern innovations include the loss of the nasal segment at the end of the nominalising suffix *-iben/-ipen*. The emergence of affrication in *tikno* ‘small’ > *cikno* predominates in the south, though the southern Balkans show a mixed region. By analogy to the preservation of initial *a-* segments, a strengthening of inherited initial segments is observed through addition of *a-* in words like *šun-* ‘to hear’ > *ašun-*, a development that is contained within the region between Ukraine in the north and northern Bulgaria and Serbia in the south, excluding the southernmost areas of the Balkans. South of the Great Divide, verbs belonging to the perfective inflection classes that had retained a perfective augment *-t-* are re-assigned to the class of verbs with an augment *-l-* (originally representing verb roots ending in vowels): *beš-t-jom* ‘I sat’ > *beš-l-jom*. Conservative forms occur occasionally in isolation in the south, especially along the Black Sea coast.

Some western innovations are contained and do not spread throughout the north, but continue eastwards, creating a kind of western-central innovation zone that is surrounded by retention zones. Two prominent cases in fact involve selection from a pair of competing Early Romani variants. The 2SG past-tense and present copula conjugation marker *-al* was probably the older historical form (going back to the 2SG oblique enclitic pronoun **te*). In Early Romani it appears to have competed with *-an*, an analogy to the 2PL marker. The form in *-al* is generalised in the western innovation zone in Germany and spreads eastwards into central Europe to include the Romani dialects of historical Habsburg Monarchy and on to

some of the dialects of Trans-Carpathian Ukraine, but leaves out the entire western periphery (Britain and Spain) as well as northern Poland and the Baltic areas. A very similar diffusion pattern is found for the predominance of *-h-* over *-s-* in grammatical paradigms and in particular in intervocalic position such the singular instrumental/sociative case endings (*leha* 'with him' vs. *lesa*). Here too, the variation appears to go back to Early Romani. Note that *s/h* alternation is found in a wide transition zone encompassing the continental side of the Adriatic and stretching all the way to Transylvania. Finnish Romani matches this western-central diffusion zone for both items, indicating a rather early development, prior to its separation from the continental dialects.

Other prominent isoglosses divide the Romani speaking landscape into further zones. Some outcomes of the western developments are contained even further and remain limited to Romani varieties spoken within the German-speaking area and neighbouring regions. These include the shortening of *anglal/angil* 'in front' to *glan/gil*, of *ame* 'we' to *me*, and of the verbs *ačh-* 'to stay' and *av-* 'to come' to *čh-* and *v-* (as examples for numerous other items affected by the process). The areas south of the Great Divide remain unaffected by these developments, but they are not replicated through the entire northern zone either. Instead, a northeastern zone emerges, with its epicentre in northern Poland, comprising the Baltic coast and North Russia and usually reaching northern Ukraine. Here, jotation appears consistently so that *ame* 'we' becomes *jame*, and the verbs *ačh-* 'to stay' and *av-* 'to come' become *jačh-* and *jav-*. A partition similar in shape emerges around analogies in the past-tense marker of the 2PL. The original *-an* prevails in the northwest as well as in a central belt connecting Germany all the way with the Romanian Black Sea Coast. The innovation centres are once again the northeastern zone, comprising Poland, the Baltics, Russia and Ukraine, where the predominant form is *-e* (by analogy to the 3PL), and the southern periphery, from

southern Romania through to the Mediterranean coast of France, where a partial analogy renders the form *-en*.

The Great Divide itself is occasionally transitional, with an intermediate central zone separating the north from the south. An illustrative example is the realisation of the word for 'horse', for which we typically find *graj* in the north, the more conservative form *grast* in the south, and an intermediate form *gra* in a central belt from the northern Adriatic to southeastern Ukraine.

Finally, a common pattern of isogloss formation separates centres from peripheries. The generalisation of the copula stem extension in *-in-* prevails within a periphery of a southern belt of dialects stretching from southern Bulgaria through to Macedonia and the northern edge of the Adriatic coast, including the dialects of southern Italy. As mentioned earlier, isolated forms in *-in-* are retained in other dialects as well. Roughly the same area is at the same time a retention zone for the verb *ov-* 'to become'. The north tends to generalise the verb *av-* in the sense of 'to become' at the expense of the older form *ov-*. Nonetheless, some instances of *ov-* remain in the transitional dialects of Slovakia, while a similar development to that carried out in the north is also found in some of the Greek dialects. In effect, then, a three-way zone division emerges, the central one being a retention zone. An area including Ukraine, Romania, and stretching all the way to eastern Austria and southern Poland serves as a retention zone for oblique forms of the definite article in *l-*, deriving in all likelihood from remote demonstrative/ pronominal oblique forms in **oles*, **ola*, **olen*.

A prominent centre-periphery split appears in the attachment of prothetic segments in *v-*, as in *udar* > *vudar* 'door' and *ušt* > *vušt* 'lip'. The historically younger form in *v-* is found in the northernmost dialects of Scandinavia, Britain and western Europe, through to Italy and Greece and the southern Black Sea area. Conservative pockets are found north of the Black

Sea coast and along the northern Adriatic, with a mixed zone stretching from southern Bulgaria to Transylvania. A more coherent conservative zone, completely lacking forms in *v-*, appears in Latvia, Lithuania and northern Poland.

The picture for lexical items in *a-* is almost a mirror image. For *angar* > *vangar* 'coal', the conservative form *angar* prevails in the centre, with the form *vangar* appearing in the Baltics (from northern Poland to Estonia), Britain and Italy. For *ařo* 'flour', the centre has conservative *ařo* in the south and jotated *jařo* in the north, while *vařo* prevails in the entire periphery belt of Finland, Britain, Italy, Greece and Crimea. With *aver* 'other', the spread zone of *vaver* is considerably wider, comprising the entire west and the Baltics as well as Greece, while *javer* appears in the zone with high jotation east of Poland, and the conservative *aver* is limited to the 'traditional' southern zone stretching from northern Bulgaria to southern Poland. By contrast, a three-way division is found for 'egg', with jotated *jaro* in the north, conservative *an(d)ro* in the south and *van(d)ro* only in the extreme southeast, covering isolated dialects of the Black Sea coast in Crimea, Greece, and the dialects of southern Italy.

Further conservative peripheries appear both in geographically marginal and in 'internal' regions. The preposition *vaš* 'for' survives in the Romani dialects of Latvia as well as in the so-called 'Central' dialects of eastern Austria, Hungary, Slovakia, southern Poland and northern Romania. Greek-derived nominal endings in *-is* and *-os* survive in the geographical margins in the Baltics (eastern Finland, Estonia, Latvia), in Britain, along the Black Sea coast from Crimea through to Bulgaria and Greece (primarily *-os*, with *-is* occurring in a smaller region in Bulgaria), as well as in the Northern Central dialects of northern Slovakia and southern Poland.

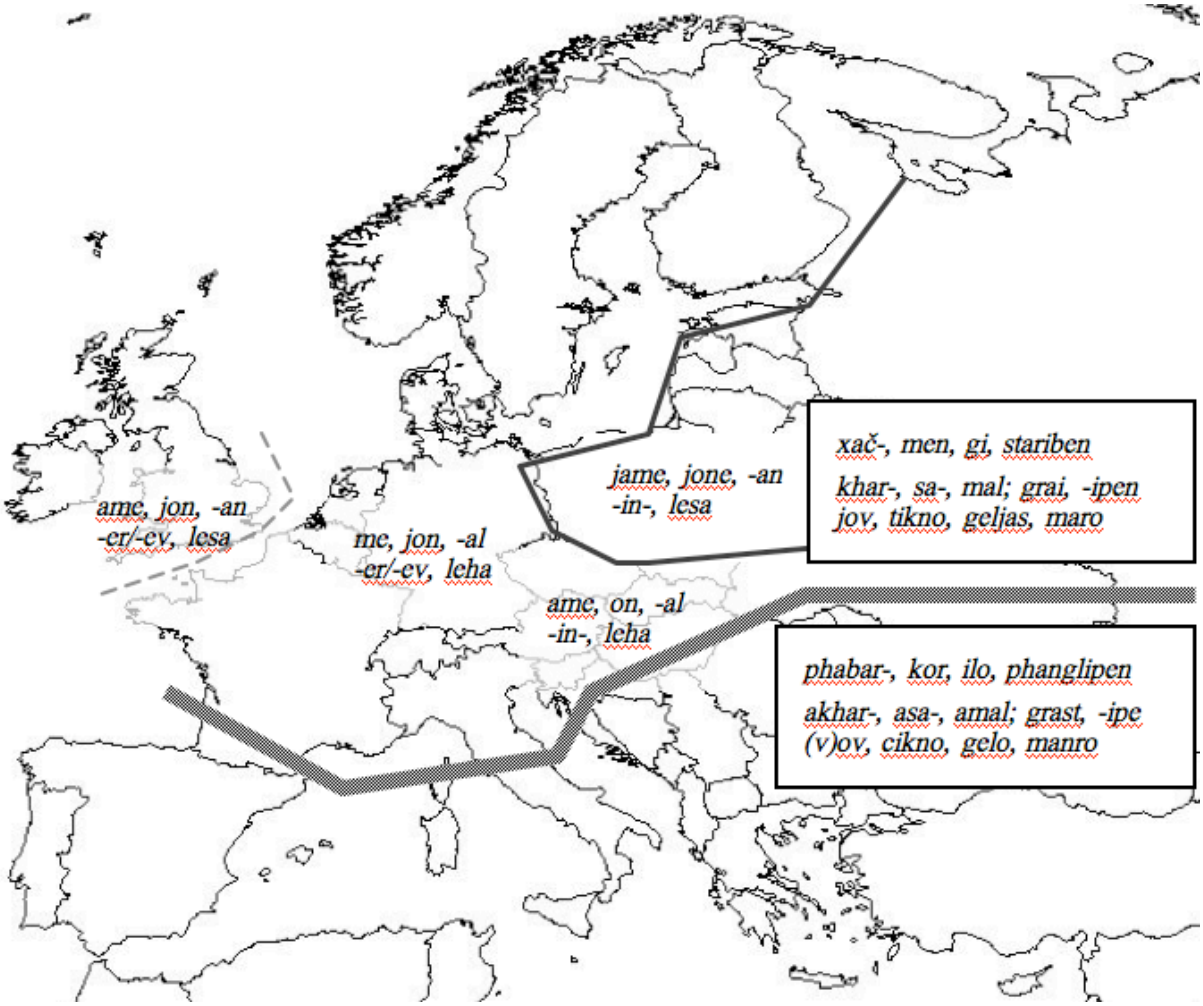
A similar conservative periphery – Britain, Spain, Italy, and the southern Balkans – shows retention of the original Early Romani demonstrative opposition set in *adava : akava* (with corresponding forms in *-o-*). The centre shows various innovation zones, where the original forms are simplified or reinforced to create opposition pairs such as *adava : dava*, *kada : kaka*, *kava : kavka* and so on. Though zones partly overlap due to the many forms that can become part of the paradigm, a rough geographical split can be identified between a zone in northern Bulgaria and Romania (*kaka*), a central zone around Hungary and Slovakia (*kada*), a northeastern zone comprising Poland and Russia (*dava : adava*) with a unique retention sub-zone in the Baltics (*kada*), a major zone stretching from the Black Sea coast to the North Sea (*kava*), and a Finnish zone (*tava*).

Finally, we find an illustrative partition into zones involving the fate of Greek-derived tense markers, incorporated into Romani as a means of adapting loan verbs to Romani inflection patterns. From the fact that here too we encounter a conservative periphery – proliferation of different forms is preserved in the dialects of present-day Greece, retention of *-isker-* also in Crimean and Zargari Romani, retention of *-isar-* both in Romania-Moldavia and in Spain, and the use of several parallel forms in Welsh Romani – indicates that Early Romani passed on a complex inventory of forms, which were later simplified in the individual dialects. The principal zones that share the same selection are the German-Finnish zone with *-er-/-ev-* (also *-ar-/-av-*), the Black Sea coast, northern Bulgaria and Greece with *-iz-*, Romania-Moldavia and adjoining regions with *-isar-* as well as contracted versions thereof, and a central-eastern zone from the Baltics and all the way down to western Bulgaria and southern Italy, with *-in-* (primarily, with additional vocalic variation in the Balkans).

Conclusions

We can try and summarise the emerging picture as follows: In relation to several prominent features in phonology, morphology and lexicon, there is a tendency toward a north-south split. The division tends to follow the older (sixteenth-seventeenth century) frontier zone between the Habsburg Monarchy and the Ottoman Empire, with innovations occurring on both sides of the divide. A southeastern zone comprising Greece and sometimes also the Black Sea coast as well as adjoining inland regions is often conservative and retains either older forms, or a greater range of variants, reflecting its position as the historical centre of diffusion. Many of the features that are specific to other zones are in fact preferences favouring one of the older variants over another, rather than structural innovations in the strict sense. In addition to this southeastern periphery, other geographically marginal zones such as Spain, Britain, Scandinavia and southern Italy also tend to show archaisms as well as non-participation in certain predominant variant selections. With respect to individual features there are of course other retention zones as well; two of the more noticeable ones, which often share retentions, are the central zone (Austria-Ukraine, or sometimes just Slovakia-southern Poland) and the Baltic zone (sometimes just limited to Latvia). Within the core (non-periphery) areas, there are further zones that tend to show coherences with respect to various features. They include the German-Finnish (northwestern) zone, the German-Hungarian (western-central) zone, the Romanian-Moldavian (Wallachian or Vlax) zone (with its further penetration into Serbia and Croatia as well as Hungary and migratory spread even beyond these regions during the past two centuries), and the Polish-Baltic (northeastern) zone.

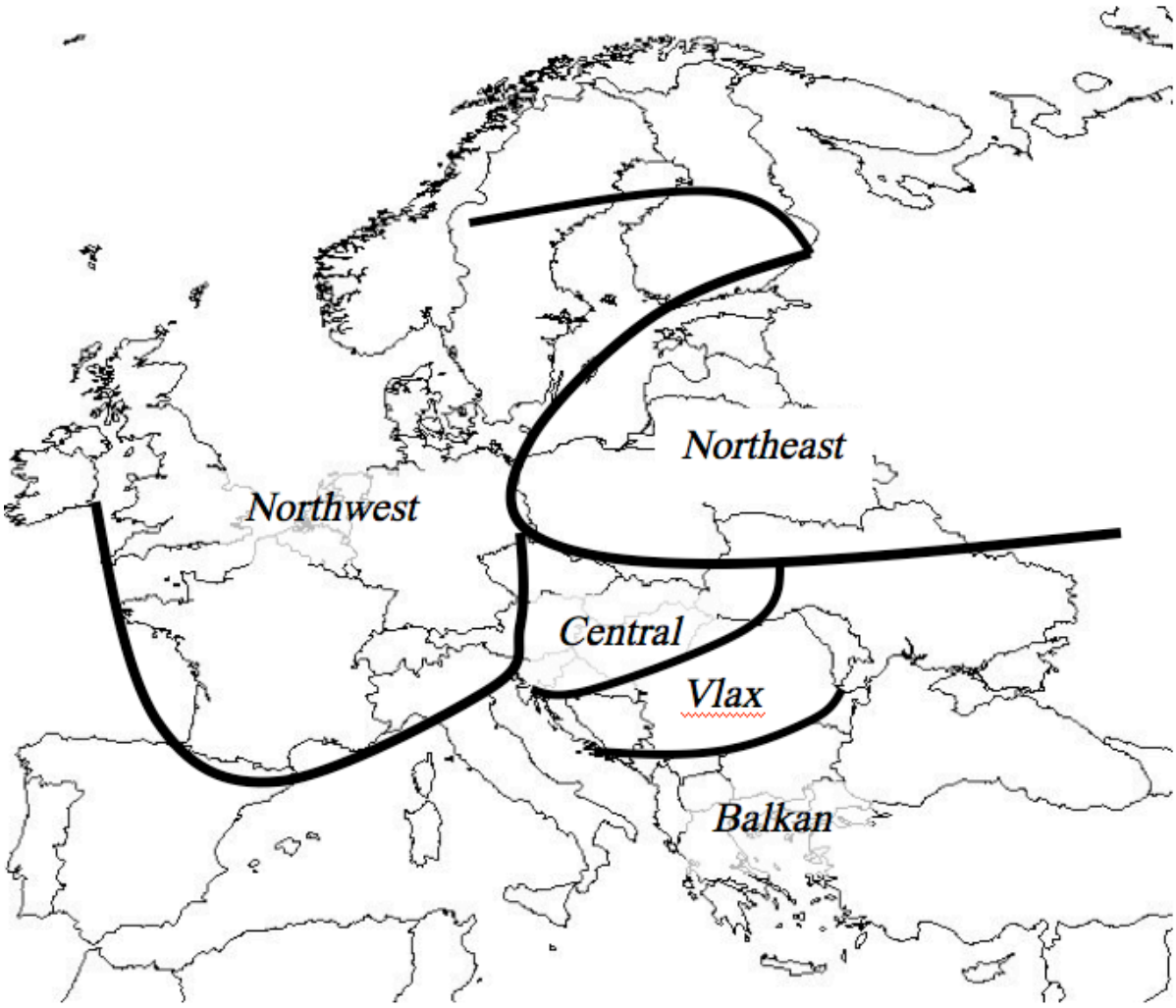
Map 1: Some major isogloss clusters in Romani



It must be emphasised once again that the participation of a particular region in a 'zone' with adjoining regions is by no means pre-determined or static. Different isoglosses show different extents of diffusion, often in different directions, reflecting in all likelihood changing patterns of social networking over time. In some instances these networking patterns will have been influenced by concrete impediments such as political boundaries and the migration of population groups away from their earlier locations. In other cases, they might reflect shifting alliances between groups and consequent shifts in the prestige centre.

The coherences in the distribution patterns of some of the more salient morpho-phonological features have inspired scholars in Romani linguistics to apply a kind of reference grid for the classification of dialects. The prevailing discourse in Romani linguistics, first formulated explicitly in Bakker & Matras (1997; see also Matras 2002 and Elšík & Matras 2006 for an in-depth enumeration of groups and features) has since recognised the following dialect 'groups': Northern, sub-divided into Northwest (German-Scandinavian), Northeast (Polish-Baltic), and British; Iberian; Central; Vlax; and Balkan. Aware of the rough nature of this classification, most authors tend to recognise sub-divisions, splitting Central, Vlax, and Balkan into northern and southern sub-branches respectively. Still, there remain a number of regions and dialects that are not easily accommodated in this reference grid, such as the Romani dialects of southern Italy or some of the Romani dialects of Ukraine, and there are numerous varieties of Romani that show combinations of features that are conventionally attributed to two distinct 'branches'.

Map 2: Conventional dialect classification grid in Romani linguistics



All this goes to show that there is no static, pre-determined 'membership' in a dialect group. Dialect groups are therefore not 'families' that prescribe an inescapable or unambiguous 'genetic' affiliation. They are, rather, terms of convenience that allow researchers to make generalisations about the distribution of structural features using an economical inventory of reference terms.

REFERENCES

Bakker, P., and Matras, Y. 1997. Introduction. In: Matras, Y., Bakker, P., and Kyuchukov, H. eds. *The typology and dialectology of Romani*. Amsterdam: Benjamins. vii-xxx.

Boretzky, N. 2007. The differentiation of the Romani dialects. *Sprachtypologie und Universalienforschung* 60, 314-336.

Boretzky, N. & Igla, B. 2004. *Kommentierter Dialektatlas des Romani*. Wiebaden: Harrassowitz.

De Goeje, M. J. 1903. *Mémoire sur les migrations des tsiganes à travers l'Asie*. Leiden: Brill.

Elšík, V., and Matras, Y. 2006. *Markedness and language change: The Romani sample*. Berlin: Mouton.

Gilliat-Smith, B.J. 1915. A report on the Gypsy tribes of North East Bulgaria. *JGLS*, new series, 9: 1-54, 65–109.

Hancock, I. F. 2008. Mind the doors! The contribution of linguistics. Paper presented at the international conference 'All Change! Recentdebates over the history and origin of the Roma/Gypsies/Travellers, University of Greenwich, June 2008.

Marsh, A. R. N. 2008. "No Promised Land" History, Historiography and the Origins of the Gypsies. Unpublished PhD thesis, University of Greenwich.

Matras, Y. 2002. *Romani: A linguistic Introduction*. Cambridge. Cambridge University Press.

Matras, Y. 2004. Romacilikanes – The Romani dialect of Parakalamos. *Romani Studies* 14: 59-109.

Matras, Y. 2005. The classification of Romani dialects: A geographic-historical perspective. In: Ambrosch, G., Halwachs, D. and Schrammel, B., eds. *General and applied Romani linguistics*. Munich: Lincom Europa. 7-26.

Turner, R. L. 1932. So-called prothetic V- and Y- in European Romani. *JGLS*, Third series, 11: 115-120.