

The classification of Romani dialects: A geographic-historical perspective

Yaron Matras

University of Manchester

1. Introduction

The prevailing approach to dialect classification in contemporary Romani linguistics is one that places ‘branches’ and the branching metaphor in the centre of the analysis. Individual dialects are regarded by most authors as belonging to one of several branches, which in historical perspective implies that they are offshoots of a proto-variety. Although there has been no attempt to date to actually reconstruct features of any one of those proto-varieties, the general understanding in the literature is that individual proto-varieties emerged either during the Byzantine period of Romani, or shortly thereafter, that is, following emigration from the southern Balkans.

This viewpoint is largely inspired by Miklosich’s (1872-1880) model, which traced the migrations of Romani groups and their successive splits from one another across Europe. It is further reinforced by an assumption that, since Romani lacks coherent and continuous territorial representation, its dialects are not subjected to a geographical diffusion pattern of innovations, and consequently do not form a geographical dialect continuum. Tailored to the indeed rather complex co-existence of different layers of Romani populations in the southern Balkans, there is, in fact, a tradition of classifying Romani dialects not by location, but by the occupation and origins of their speakers (see Paspatis 1870, Gilliat-Smith 1915), and so as ‘insular’ rather than ‘territorial’ (Boretzky 1998).

The continuing prevalence of the branching model might also be attributed to the absence, so far, of any modern and systematic study of cross-dialectal variation in Romani, which might have allowed scholars to examine the position of individual features

in individual dialects in both historical and geographical perspective. The present paper offers to fill part of this gap by arguing for a geographical diffusion model that can account for the major isoglosses separating Romani dialects. According to this model, relations between dialects are not absolute, based on 'genetic' criteria, but relative: dialects are more closely, or more remotely, related to other dialects, depending on the number of relevant features that they share. The structural features that distinguish dialects are a result of processes of change and innovation that spread from one community to another. The outcome of these changes can be plotted on a map in the form of isoglosses. Dialects thus form a geographical continuum which reflects the historical spread of structural innovations (as well as the clustering of archaisms) in time and space. The closer two dialects are on the map, the greater the likelihood that they will share structural features; exceptions are those dialects that are positioned close to a bundle of several isoglosses. While it will be impossible to give thorough consideration to the historical developments in this article (the reader is referred to Matras 2002 for a more elaborate discussion), the point of the paper is to highlight principal isoglosses.

These principal isoglosses appear to have emerged during the period of settlement that followed emigration from the Balkans, that is, between the fifteenth and seventeenth centuries. Whether or not a dialect is likely to participate in any specific isogloss development is therefore dependent upon the geographical location of that dialect *during the relevant period*. The model therefore has to consider the original location of migratory dialects that have been displaced after the period of Romani settlement. If due consideration is given to changes of location, the emerging picture in respect of a series of developments is one of largely coherent *diffusion spaces*. I propose to define the relationships between Romani dialects in terms of the individual diffusion spaces to which they belong.

2. Models of Romani dialect classification

Dialect classification in Romani has its roots in Miklosich's (1872-1880) comparative survey and historical discussion. Miklosich's classification was based on a reconstruction of the migration routes of the Romani groups that had left the southern Balkans. The principal reference features, however, were not innovations within the internal, inherited component, but rather the successive layers of loan vocabulary. The result was a reconstruction of the branching and sub-branching of groups from several main waves of migration, a grid that later inspired Romani dialectologists to postulate several dialect branches, seen as a 'genetic' split.

A different kind of approach to dialect classification in Romani was taken by Gilliat-Smith (1915), focusing on the dialects of northeastern Bulgaria. Gilliat-Smith recognised that in the area under his investigation, dialects belonging to different branches overlap geographically. The distinction between 'settled' and 'nomadic' dialects had already been introduced for the Balkans by Paspati (1870). Gilliat-Smith adopted the term *vlax*, used by immigrant (mainly Orthodox, and nomadic) Rom originating from Wallachia, and contrasted it with the *non-vlax* (mainly Muslim, and settled) Romani populations. Due to the frequent presence of immigrant communities speaking Vlach Romani dialects in other parts of Europe, for a while this distinction was adopted as a kind of 'basic' dialect division within Romani. Occasionally, authors may still characterise a particular dialect as being 'non-Vlach', even if it is spoken in a remote location from the Vlach dialects, or even if it has never been argued to be a Vlach dialect, and so does not really need to be presented explicitly as non-Vlach.

Gradually, a division into several dialect groups of equal ranking has emerged, which has become a popular reference grid in work on Romani linguistics during the 1990s (cf. Bakker & Matras 1997, Elšík 2000). This division recognises the following branches:

- a) Vlach, centred around the historical Wallachian and Transylvanian regions, with out-migrants in various regions throughout Europe and beyond;

- b) Central, with a northern sub-division (Northern Central) in southern Poland, northern Slovakia, and Transcarpathian Ukraine, and a southern sub-division (Southern Central) in southern Slovakia, Hungary, southeastern Austria and northern Slovenia;
- c) Balkan, including the Black Sea coast dialects, occasionally sub-divided into a ‘default’ Balkan dialect – Southern Balkan I in Boretzky’s (1999a) terminology – and a distinct sub-group based in northeastern Bulgaria and Macedonia – called Southern Balkan II (Boretzky 2000) or the Bugurdži-Kalajdži-Drindari group.
- d) More controversial are the status and affiliation of the dialects of western and northern Europe, including southern Italy and the Iberian peninsula. Bakker (1999), following other suggestions in the literature, had grouped them together under the heading of a so-called ‘Northern’ branch. An alternative is to define separate North-eastern (Baltic) and Northwestern (German-Scandinavian) groups, and to treat the remaining dialects as isolates (see Matras 2002, Ch. 9).

Note that none of these divisions is defined by a strict set of specific structural features. The division into groups is usually based on impressions of a series of shared features and so of some degree of internal coherence, often accompanied by geographical proximity. Attempts to enumerate shared features have taken for granted the branch affiliation of individual dialects, and worked their way inwards to take an inventory of features that occur within the nomenclature (see Boretzky 1999a, 1999b, Bakker 1999). The problem is, that membership in the nomenclature itself is to a considerable degree arbitrary, or at least based on non-systematic criteria. There is also the unsatisfactory circumstance of classificatory inconsistency: while some groups may be defined in relation to a particular structure, that structure may be irrelevant to the characterisation of another group.

The notion surrounding the ‘genetic’ classification is that a larger inventory of shared features is a token of an historically coherent population group, which, if spread over a larger territory, will have spread as a result of migration. If this group shares a limited inventory of features with another group, then this is taken to represent earlier

ties with that group, prior to the break-away through migration of one of the populations. Thus, Boretzky (1999a) suggests that the southern Central dialects differ from the northern Central dialects due to a late immigration from the south into their present location, which also explains the fact that they share a number of features with the Balkan dialects. The plotting of feature inventories on regional maps, which Boretzky (1999a, 1999b) has introduced, has so far largely received an interpretation within this ‘genetic’ model, and little attention has been paid to the possibility that may have been adopted by speaker populations in situ. Indeed, most feature inventories that have been taken in comparison between dialect sub-groups fail to undertake a systematic differentiation between innovations and archaisms. As a result, it is not always apparent which of the features are simply common retentions from Early Romani, and which can be taken to represent distinct changes that occurred within the speech of a particular close-knit network of speakers, and hence are relevant for classificatory purposes.

3. Diffusion in geographical space

The point of departure of the geographical diffusion model (outlined in Matras 2002, Chapter 9) is that some differences among dialects are more, and others are less, relevant to classification. Those that are more relevant are the result of structural changes that gave rise to innovations which spread through large geographical spaces. Features that are less relevant are occasional archaisms that form no coherent pattern, and where preservation of a feature is due, in all likelihood, to relative isolation of the speakers during the time in which, in other dialects, change occurred. Less relevant are also innovations that are widespread and common, and appear in individual dialects irrespective of location. Shared innovations tell us something about the connections between the speaker populations of the dialects in which they occur, while common innovations do not.

For example, the raising of the vowel in the conjunction *te>ti* is found locally in a number of varieties, such as German and Austrian Sinti, as well as Sepečī. It is not a

development that has enjoyed widescale diffusion, and it does not help determine the relation of a dialect to other dialects. Similarly, the loss of overt, consonantal perfective markers, as in *kerdom* > *kerom*, is found in various locations, among them some Sinti dialects, and some dialects of northeastern Bulgaria (Bugurdži). Likewise, this is not a principal isogloss, since it results from separate local developments that happen to have led to a similar outcome. It is a *common* development, rather than a *shared* development – a development that the dialects underwent together. Some contact-related developments may also be common, as they depend not on the shared change within Romani, but on the features that happen to be common to the various contact languages. The productivity of morphological causative formations is one such feature, triggered by contact with the typologically somewhat similar languages Turkish (in Sepeči or the Ajia Varvara dialect), and Hungarian (in Eastern Slovak Romani).

By contrast, the disappearance of the older, gender-marked form of the intransitive perfective *gelo*, *geli*, and its substitution by a form that copies that inflectional patterns of transitive verbs, *geljas* (*gejas*, *geja*), appears in a coherent area, in a rather systematic way, and so it can be regarded as a *shared* development. It is a development that divides the Romani-speaking landscape into two: an area in which the innovation has been adopted, which I call below a *diffusion space*, and an area which has resisted the innovation.

Following the diffusion model, we are interested in the formation of such larger *diffusion spaces*. Our assumption is that diffusion will be indicative of the contacts that existed between the speaker populations at the time just following the initiation of the change. Not all changes emerged at the same time, however, and so not all changes may have enjoyed the same pattern of contacts, and spread in the same way. Moreover, some changes might have been adopted more readily by neighbouring groups, while it is possible that others were seen as markers of an identity that was not a target of imitation, or as otherwise less prestigious. For this reason, different features may show distinct diffusion spaces.

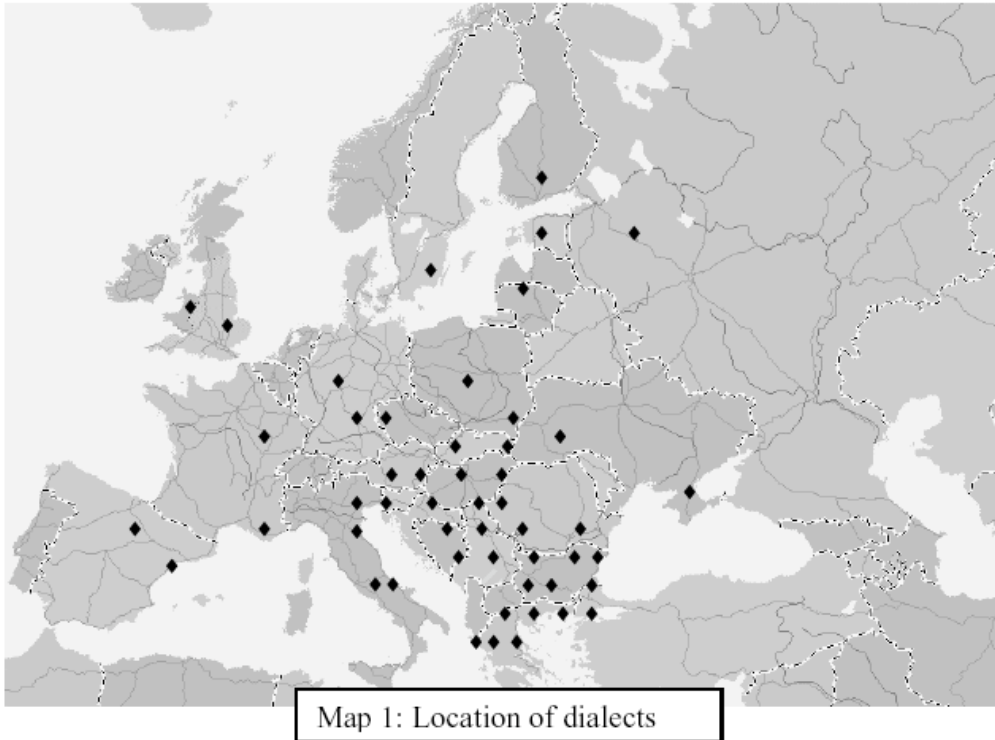
The position of a dialect on the dialect continuum is best defined in relation to a series of isoglosses marking out the diffusion spaces for individual changes. If we are looking for a way to divide the dialects of Romani into groups, then the sensible way to do this is therefore to search for *bundles* of isoglosses, that is for overlap among diffusion spaces. Developments that cluster are more easily indicative of patterns of settlements and historical contacts, and so of the coherence of speaker populations during the period at which the relevant structural change took place.

In the following, I will concentrate on the clustering of several isoglosses that form larger diffusion spaces within the Romani-speaking landscape. The data which are taken into consideration derive from the Romani Morpho-Syntactic Database (RMS) developed by Viktor Elšík & Yaron Matras at the University of Manchester. The database documents variation in Romani, drawing on both published sources and on fieldwork conducted in Romani-speaking communities throughout Europe.¹ A characterisation of the major dialects and the published sources is provided in Matras (2002, Chapter 2), and will not be repeated here. Map 1 illustrates the distribution of locations of individual dialects taken into consideration in the present study.

4. The Great Divide

If we concentrate our attention on primary isoglosses, or those that represent large-scale shared innovations, and if we examine the geographical patterns that they form, then it is possible to identify two primary centres of innovation in Romani. The first is in the northwest, and its centre is in or around Germany. The second is in the southeast, where two distinct distribution patterns can be recognised: The first covers the entire Balkans. The second is more specific to Transylvania/Wallachia, but often influences the Balkans, especially the Black Sea coast, sub-dividing the southern Balkans into an eastern and a western zone, as well as the Transcarpathian areas to the northwest of it.

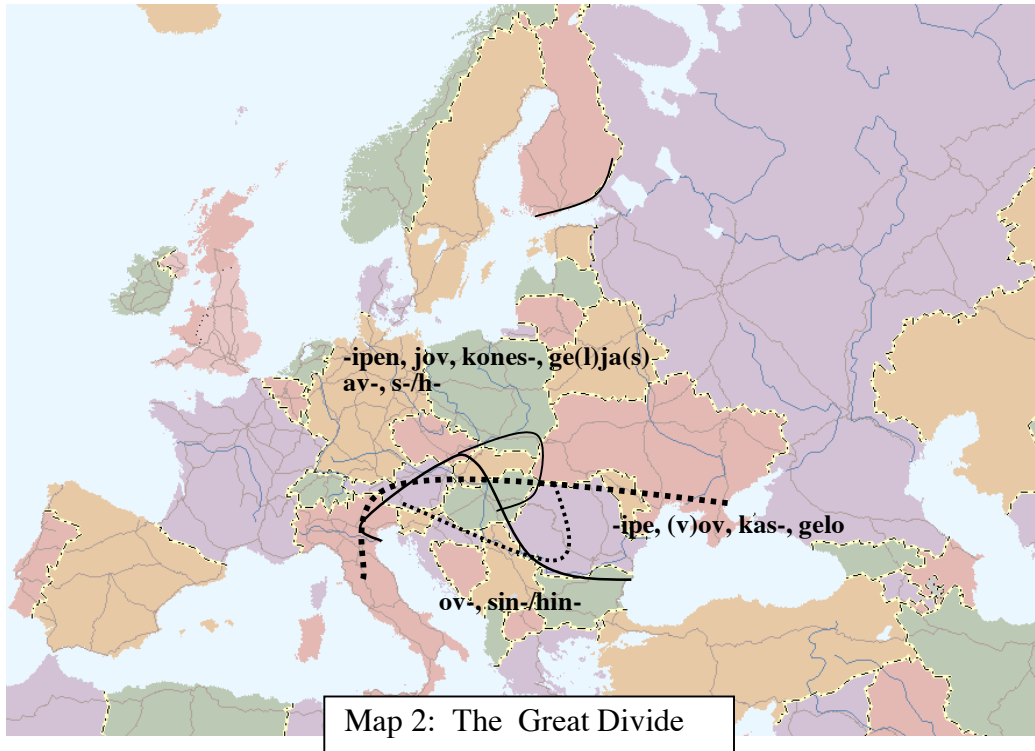
¹ I acknowledge support for the project from the Arts and Humanities Research Board and the Open Society Institute.



The two major centres are separated by the Great Divide – a bundle of isoglosses cutting through central Europe in a transition area stretching between the line southeastern Austria-Hungary-Romania in the north, and the line Slovenia-Croatia-Vojvodina and the Danube in the south (Map 2).

Innovations spread on each of the two sides of the Great Divide, but each side's innovations are contained by the Divide and do not spread across it. To the north we find retention of the nasal in the nominal suffix *-ipen/-iben*, which is lost in the south, resulting in *-ipe/-ibe*. A northern innovation by contrast is the prothetic insertion of *j-* in a series of words beginning in a vowel, especially in *a-* (*aro* > *jaro* 'egg'), and characteristically in the form of the third person pronoun (*ov, oj, on* > *jov, joj, jon*). This prothetic insertion is not typically found in the south. The north also shows a series of analogous morphological renewals: The oblique form of the interrogative 'who' is restructured throughout the north to *kon-es*, based on the nominative *kon*, while in the south we find retention of the older (and irregular) form *kas*. The 3.SG of past-tense in-

transitive verbs is assimilated into the transitive paradigm in the north, where we find *geljas* ‘he, she went’ (or *gejas*, or *geja*), by analogy to *kerdjas* (*kerdja*) ‘he, she did’, while in the south we find the older, adjectival forms *gelo* ‘he went’, *geli* ‘she went’.



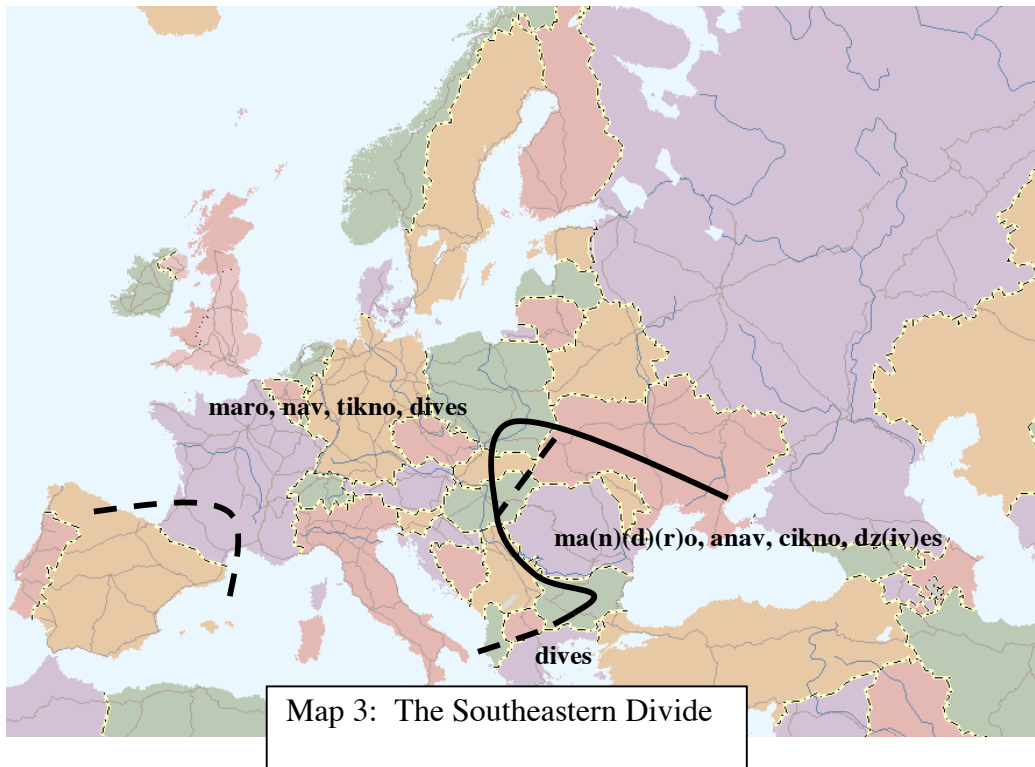
These developments are represented by the dotted line, which also indicates a transitional zone in which the latter phenomenon – *gelo* > *geljas* – has advanced only more recently, and where the two forms are often found alongside one another, and in some dialects in functional distribution (see e.g. Matras 1995). The full line represents two further developments that characterise the Great Divide. In the north, the subjunctive and future form of the copula ‘to be’ is renewed to *av-*, modelled on the verb ‘to come’, whereas in the south the original form *ov-* ‘to become’ is retained. Also in the north, the copula forms tend not to show any insertions in *-in-*, although these are found within a transitional zone in and around Slovakia, as well as in Finnish Romani (*hin* ‘is’), while

in the south, *-in-* appears more regularly, either confined to some tenses (*sine* ‘was’), or throughout the paradigm (*sinom* ‘I am’ *sinan* ‘you are’).

The isogloss bundle of the Great Divide is partly strengthened by a southeastern cluster (Map 3). To the east, this line connects the dialects of the Black Sea coast region (cutting through Bulgaria) with the Carpathian region, stretching in some features as far north as Slovakia and southern Poland. In its centre, this zone contains the northern Vlax dialects, which appear to be the source of many of the innovations that characterise the eastern side of the line.

An important eastern innovation is the affrication of Early Romani palatalised dentals **t’* and **d’*, as in *cikno* ‘small’ and *dzes* or *zis* (from **dzives*) ‘day’, against the loss of palatalisation in the western side (*tikno*, *dives*). Affrication is by far more widespread in *cikno*, which reaches both East Slovak Romani in the north, and Epiros in the south, while both these regions show de-palatalisation in *dives*. Another eastern innovation is the prothesis of *a-* in many lexemes, as in *anav* ‘name’ (against retention of *nav* in the west). This innovation reaches northern Greece in the south, but not eastern Slovakia in the north. The Vlax dialects continue the development even further to include a larger number of words, such as *abijav* ‘wedding’.

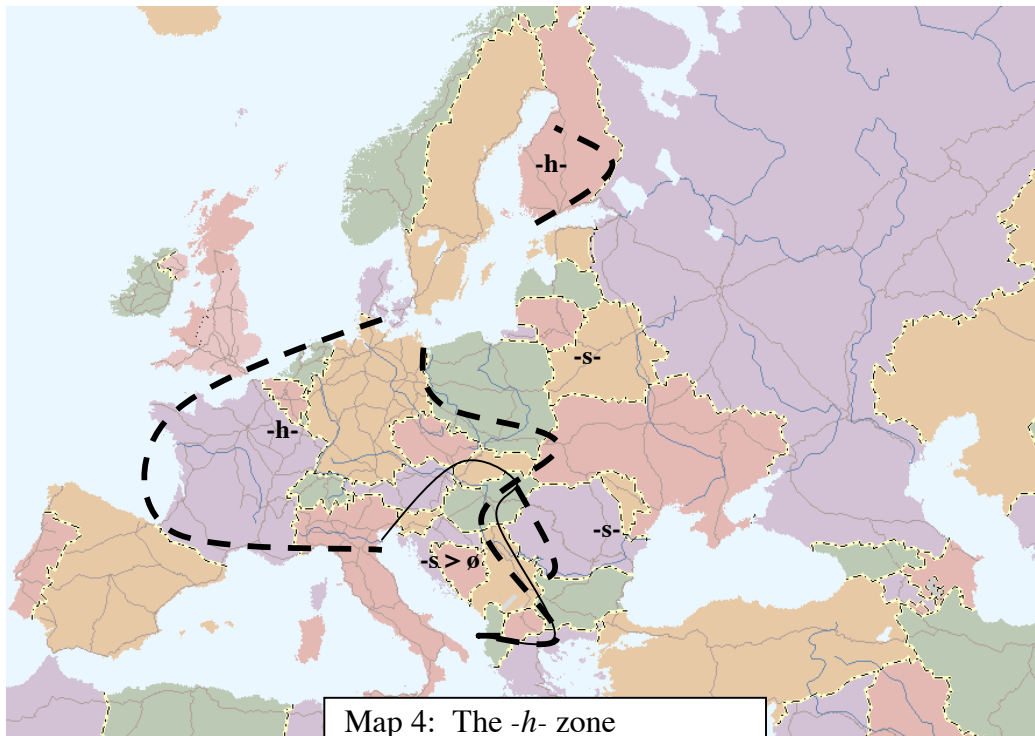
The east is on the other hand conservative in its retention of a consonant cluster in words such as *mandro* (*mando*, *manro*, *marno*) ‘bread’ (but note *maro* in northern Bulgaria), which in the west is simplified to *maro*. East Slovak Romani is situated in a transitional area in the north, and has a cluster *-ndr-* in some relevant words (*jandro* ‘flour’), and a simple *-r-* in others (*maro* ‘bread’). The dialects of Iberia form a relic area, in which occasionally both palatalised dentals and *-nr-* clusters are found; thus Catalanian Romani has *dzives* ‘day’, *manro* ‘bread’ (Ackerley 1914). The Vulcanius list of 1597, which documents vocabulary presumed to have been collected in France, has *yanre* ‘eggs’, showing the conservative form at an early date, and in the geographical periphery of the *nr>r* zone.



There are numerous other innovations based in the centre of the southeastern zone depicted on Map 3. Most are specific to the Vlax dialects, which are in the middle of the zone. Some, however, spread from the Vlax group into neighbouring groups. The Northern Vlax dialects of Transylvania share the formation of demonstrative stems in *kad-* with both East Slovak Romani to the north, and some of the northern Bulgarian dialects of the Drindari-Bugurdži group, to the southeast. With the latter they also share demonstratives in *kak-*. An archaism of the zone is the preservation of loan verb adaptation markers in *-is-* and/or *-iz-* from the Early Romani inventory. The second, *-iz-*, is typical of the Black Sea coast dialects, while *-is-* (in *-is-ar-*) appears in the Vlax group, as well as in the relic areas of Iberia and Wales (see discussion below).

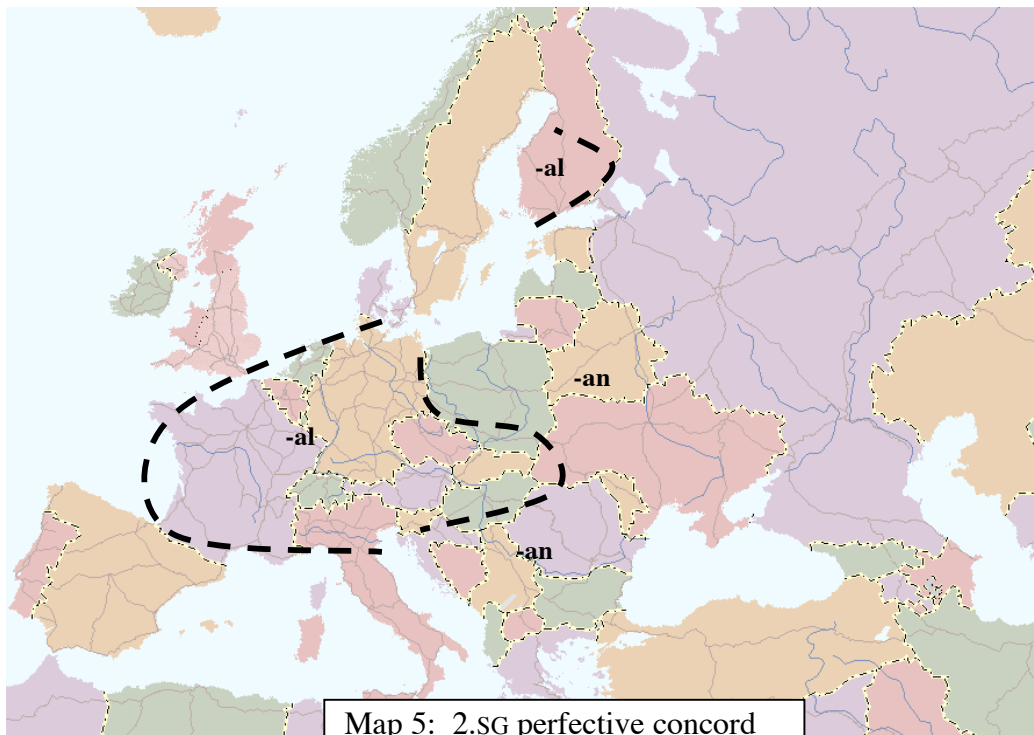
5. The central European or *h*-zone

A salient structural variant in Romani is the alternation of *-s-* and *-h-* in grammatical paradigms (see discussion in Matras 2002, Ch. 4). The alternation itself appears to be an archaism, as is the retention of two full sets of copula forms, in *-s-* and in *-h-*, in a number of dialects (Arli, Italian Sinti). Other dialects have retained just one set, generalising either *-s-* or *-h-*, or combining *-h-* in the third person with *-s-* elsewhere. The presence of *-h-* in any of the copula forms always coincides with the generalisation of *-h-* in intervocalic position in grammatical markers (*dž-aha* ‘we go’, *leha* ‘with him’, *kerelahi* ‘he used to do’, etc.). The spread of *-h-* to determiners and interrogatives generally coincides with the generalisation of *-h-* in the copula. The core of the process is thus the selection of forms in *-h-* in intervocalic position and in the third person copula (*hi* or *hin*). This occurs in a coherent geographical zone in central Europe, as well as Finland, as shown in Map 4.

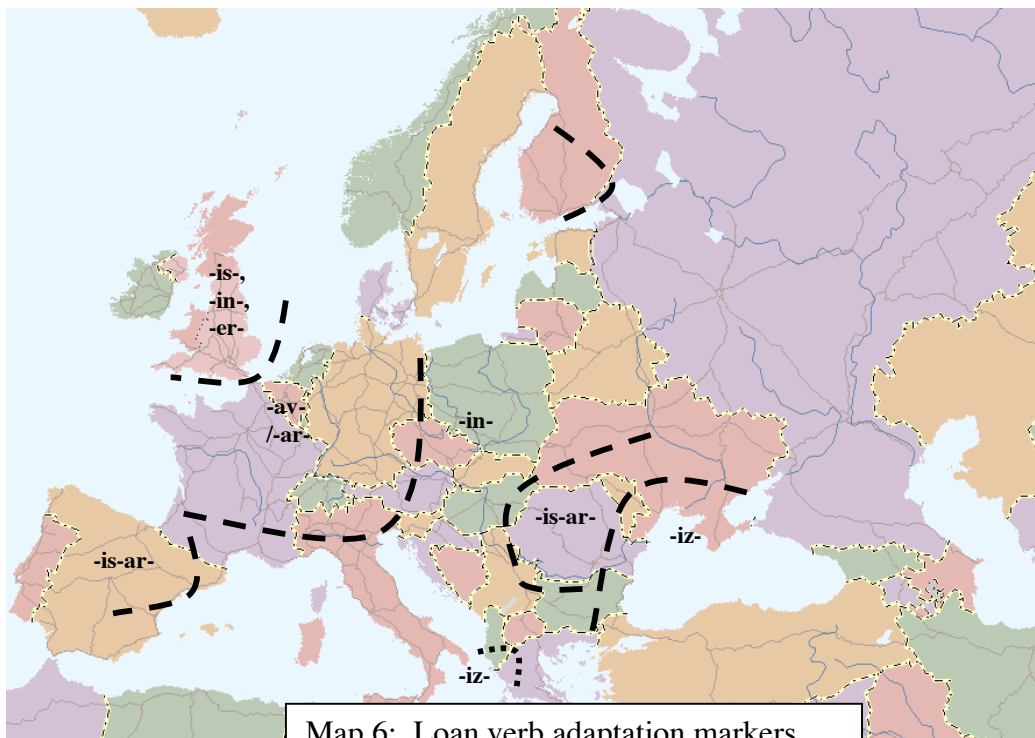


The map illustrates that *-h-*, like other features discussed in the previous section, is not a ‘genetic’ feature acquired prior to settlement outside the Balkans, but a feature that emerged in situ. Unlike some of the other features discussed so far, however, it is less likely to have been a plain sound innovation, but rather seems to have resulted from the selection and generalisation of an existing option, one that has disappeared outside the zone. Dialects with intervocalic and copula *-h-* are the Sinti, Finnish, Central, Doljenski (Slovene-Croatian), Gurbet (south-western Vlax), some Arli varieties, transitional varieties of Vlax and the Central dialects such as Cerhari and Gurvari, as well as fringe dialects of the zone, such as that of Serres in northern Greece. The process has also partly advanced to include some of the Kelderaš varieties in Serbia and Transylvania.

A more straightforward sound development is the loss of *-s* in final position in a zone on the eastern Adriatic coast. Dialects with loss of final *-s* include Doljenski, the Southern Central group, southwestern Vlax and Arli varieties.



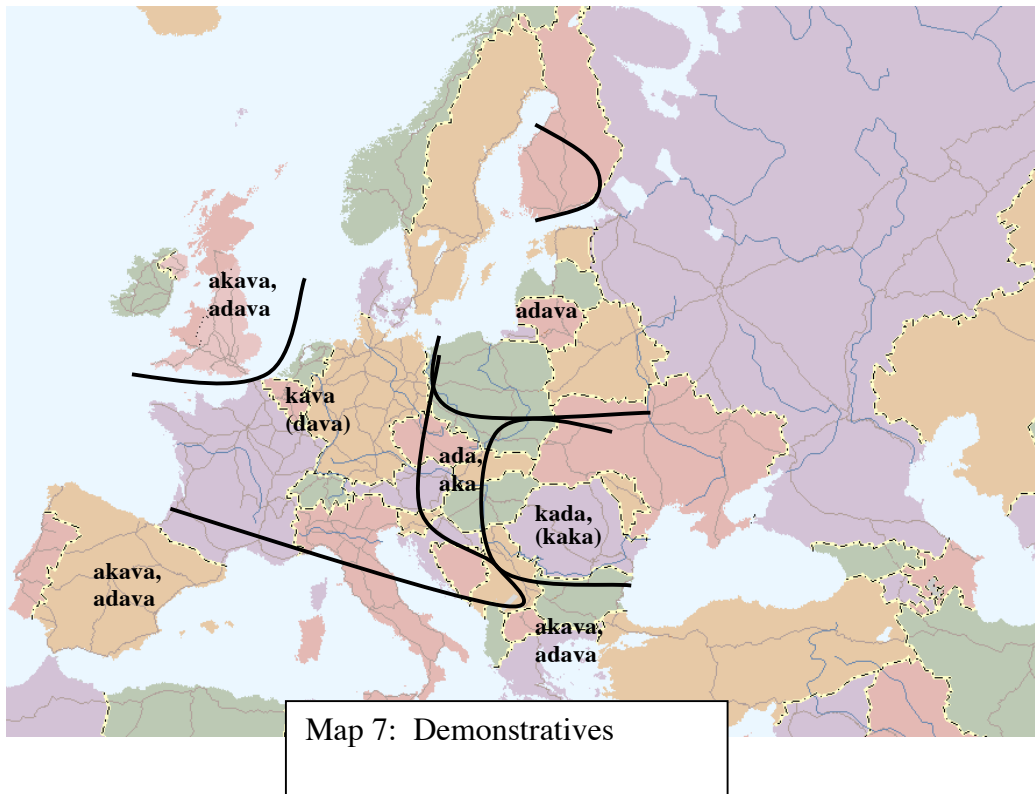
Remarkably similar to the isogloss representing *-h-* in internal (grammatical) position is another zone, within which a differentiation is found between the perfective concord markers of the 2.SG (*-al*) and the 2.PL (*-an*), both also used in the present copula. Outside the zone, both persons merge in *-an* (Map 5). Although *-al* is the more conservative form (going back to the oblique pronominal clitic **-te*; see Matras 2002, Ch. 6), the innovative analogous formation in *-an* is widespread and not contained in any specific area. This suggests that here too, we are dealing with the selection of one out of two variants that were already available during the Early Romani period, and so with the generalisation of an option. We thus have, in the intervocalic *-h-* zone and the *-al* zone, more-or-less overlapping isoglosses, representing a similar historical development of option selection and generalisation, in two different components of structure.



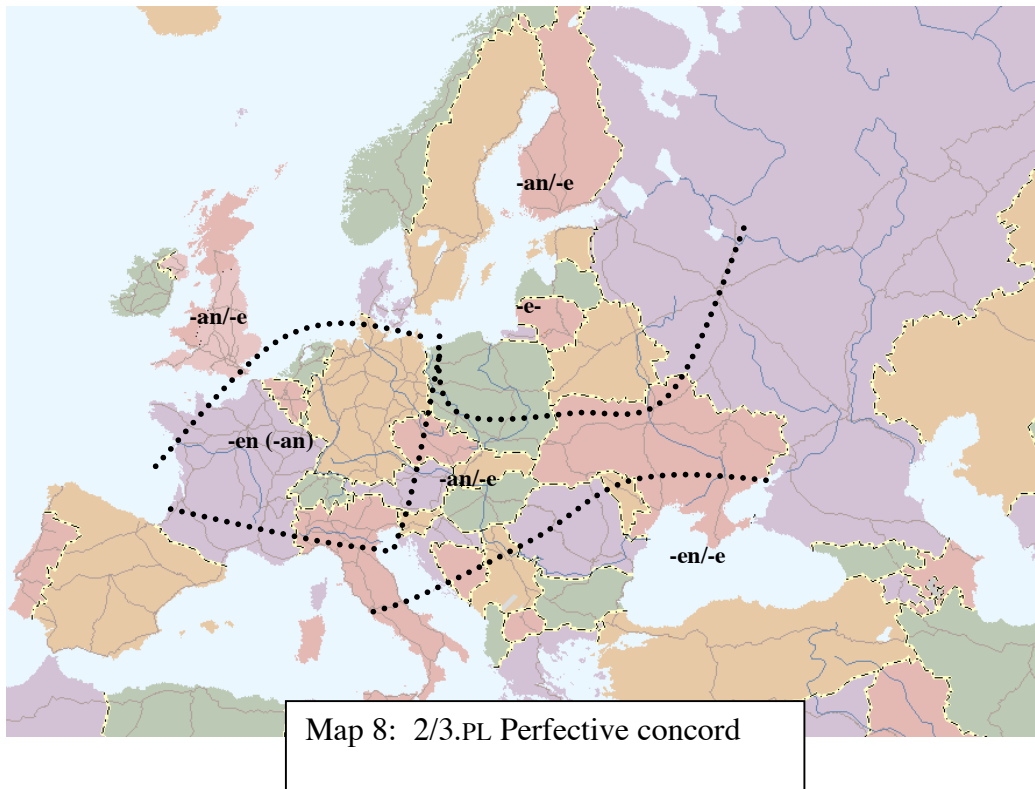
6. Paradigm renewal and the consensus reference grid

So far, the isoglosses discussed here overlap only partly with the reference grid or so-called ‘consensus classification’ into the Balkan, Vlax, Central and Northern (Northwestern and Northeastern) branches. The reference grid appears somewhat more relevant, however, when we consider the clustering of isoglosses that represent the restructuring of a number of complex morphological paradigms, notably demonstratives, loan verb adaptation markers, and further analogies in the set of perfective concord markers.

Map 6 shows the geographical distribution of markers used in the verb to indicate borrowings from contact languages (loan verb adaptation markers). These are either Greek-derived, or derived from inherited transitivity markers, or from combinations of both. The complete inventory was originally shared by all dialects, then simplified, with individual dialects retaining just one or two forms (see Matras 2002, Ch. 6). We find *-iz-* around the Black Sea coast, as well as in northwestern Greece (Epiros); *-is-* (in *-is-ar-*) in the Vlax dialects; *-in-* in a belt comprising the Baltic dialects (Northeastern group), the Central dialects, and the western Balkan dialects; and *-av-* or *-ar-* in the Sinti and Finnish groups (Northwestern). The now extinct dialects of Britain and Iberia appear to have retained a more conservative pattern, with several different forms attested in Welsh Romani, and at least *-is-ar-* attested in the fragmented documentation of Iberian Romani.



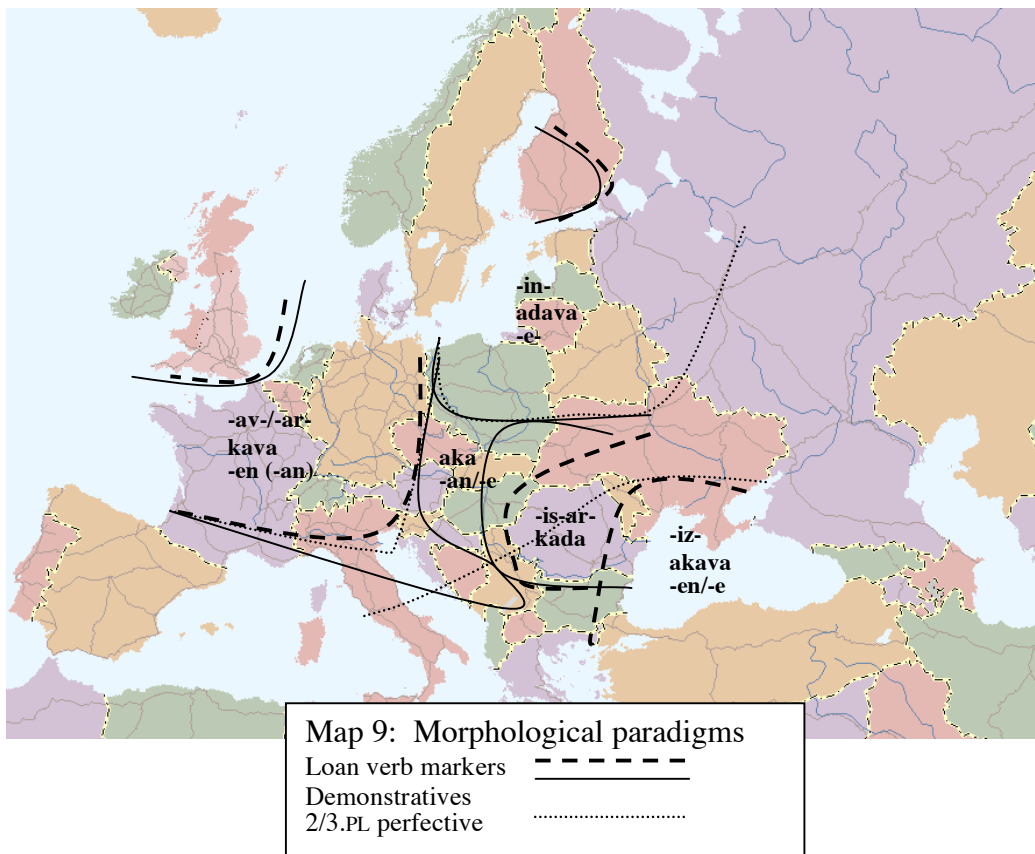
Map 7 shows the distribution of demonstrative forms. The older forms *akava* and *adava* are found in the Balkans as well as the southern and western peripheries (Iberia, Italy, and Britain). In the southeast we find a renewed form in *kada* (also *kaka*), centred around northern Vlax in Transylvania, but spreading as far as East Slovak Romani in the north, and the Drindari-Bugurdži dialects of the Bulgarian Black Sea coast in the south. In a region of central Europe, shortened forms in *ada*, *aka* are common. Western dialects have reduced forms in *kava* (also *dava*), which is found in Sinti, Finnish Romani, and some of the Adriatic coast dialects, while the Northeastern group has *adava* (but has lost **akava*).



Finally, Map 8 represents changes in the personal endings of the second and third persons plural in the past (perfective) tense. The original forms were 2.PL *-an* and 3.PL *-e*. In the Balkans, there is partial assimilation, with 2.PL *-en* partly copying 3.PL *-e*. In the Northeastern group, both forms merge in *-e*, while in Sinti both merge as either *-en* or *-an*.

The three features combine in some areas to form a dense bundle of isoglosses (Map 9). Thus, Britain as well as the Iberian peninsula are separate zones, respectively, often showing features in common with the southeastern Balkans. We might therefore designate them as relic areas in relation to the renewal of morphological paradigms. Another cluster contains the Sinti group of Germany and France, as well as Finnish Romani, together forming the Northwestern group. The Northeastern group is defined by a somewhat weaker bundle, comprising the isoglosses for demonstratives and perfective concord markers. Elsewhere, there is a continuum of dividing isoglosses, which tend to

form a projection from the Black Sea coast area and outwards toward the northwest. They allow a gradient separation of the Black Sea coast dialects (loan verbs in *-iz-*), the Balkan area (perfective concord in *-en/-e*), Vlax (loan verbs in *-is-ar*), and the northern Balkans and eastern central Europe (demonstratives in *kad-*).



The question arises why this particular set of features tends to cluster in a way that is so reminiscent of the so-called ‘consensus’ division into groups (taking into consideration that this group classification was not constructed systematically, with reference to any concrete set of features). There are two answers. The first pertains to the linguist’s motivation to draw, even if just intuitively, on morphological paradigms when attempting a classification: paradigms appear regularly, and so are more salient than individual

phonological, lexical or less regular morphological features. The second concerns the tendency for morphological developments to overlap. The renewal of paradigms is a complex process, which involves both a shared process of collapse, then a shared selection of analogy or other target to take over the functions of the lost member of the set. The chances of two different speaker communities following the same pattern in both steps are ever so greater if the two communities share prolonged and intensive ties, which make them a more coherent speaker population overall. The renewal of morphological paradigms is thus a reflection of the coherence of the groups in the period following settlement. It is no coincidence that this group-internal coherence is stronger the farther we look to the west, and away from the Balkans. For it is here, in the regions that are more remote from the centre of Romani population diffusion in the Balkans, that Romani populations became more isolated, and less dense (see already Matras 1999), and so less susceptible to movements and influences from neighbouring Romani communities.

7. Romani dialects in diffusion spaces

The point of departure of the present discussion was that Romani dialects are best classified in relation to their participation in diffusion spaces or isoglosses. The maps show us the diffusion spaces; I now take a two-way approach to interpreting the maps: first by phenomenon, then by dialect groups. Analysing the phenomena requires the postulation of a sequence of changes. Since no documentation of Romani exists until well into the period of settlement, and even this documentation, until the eighteenth century, is extremely fragmented, the only chronology of changes that can be postulated is an internal linguistic one, or *relative* chronology.

We begin with the largest diffusion space, that comprising all dialects north of the Great Divide. There are several distinct patterns here. The prothesis of *j-* is the most prominent characteristic of the dialects of the north and west of Europe, as the Divide cuts through central Europe (Map 2). It is one of the few innovations that are shared by

all dialects north of the Divide, including those of Britain and Iberia. Similarly consistent is the renewal in *kones-*. The renewal in *geljas* has continued to advance south during the past century; this can be seen from the distribution of the older form *gelo* among older speakers or in older sources, and sometimes for specially marked functions, in the transition zone. Fragmented documentation for Iberia does not allow to ascertain the full extent of diffusion of these latter two forms, however. But the wide distribution of the innovations suggests that they appeared at an early period, following, for sure, the emigration of the groups from the Balkans into central Europe, but perhaps prior to their further dispersion, or at least prior to the abandonment of close contacts between them.

A further diffusion space connects western-central Europe with the Adriatic coast. The most prominent isogloss representing this space is the simplification of the cluster in *maro* (Map 3), which however does not reach Iberia, nor does it reach its full extent in the East Slovak and southern Polish regions, and so it can be said to have followed the *j*-prothesis chronologically. This diffusion space is mirrored on the eastern side by the spread of innovations from Transylvania and Wallachia into neighbouring regions, best represented by the affrication in *cikno*.

More contained is the generalisation of intervocalic *-h-* (Map 4). It follows a similar pattern, but excludes Britain, Iberia, the southernmost areas of France, Italy, and in the east the Polish-Baltic-Northrussian continuum. This, and the even more contained selection of *-al* (Map 5), testify to the continuing innovation force centred in western Europe, with Germany at its core. The declining outwards projection of this centre might be taken to indicate younger processes, dating to a period in which contacts with surrounding populations had already weakened.

Map 6 (the distribution of loan verb adaptation markers) represents yet a later stage. Here, the innovation centre in Germany is disconnected even from the adjoining central European regions to the east of it, while on the other hand the cohesion between central Europe and the Adriatic coast (western Balkans) remains. The east, in turn, is split, with the Black Sea coast constituting a separate entity. The presence of several relic areas throughout the fringes also testifies to the more recent date of the change, compared to

those shown on the earlier maps. There are some similarities with the pattern pictured in Map 7 (demonstratives): Here too, the fringes constitute relic areas, and here too, Germany is separated from central-eastern Europe, though it is joined with the Adriatic coast region. The principal difference is the absence of cohesion between the Baltic area and central Europe. This indicates further coherence within smaller groups, and the emergence of regional diffusion spaces.

Noteworthy throughout is the coherence of Romani dialects in Germany and Finland, representing close contacts prior to the decline of Romani in western Scandinavia, and perhaps prior to the migration into Finland itself. Map 8, however, which shows the developments in the 2.SG perfective marker, represents a final stage in the build-up of present-day dialect groups, showing the separation of Germany from Scandinavia, and further supporting the separation of the Baltic area, seen already in Map 7.

Turning now to the relevance of diffusion spaces toward a classification of dialect groups, we face the challenge of integrating whatever evidence exists of shared migration patterns. Romani settlement in the north and west of Europe is generally less dense than in the centre and southeast. Connections among the dialects of the north and west – those often termed ‘Northern’ in some of the literature – are consequently more loose, as speaker populations will have had less intensive contacts with one another. Most shared features characterising the north are archaisms, testifying to the limited innovation potential that is shared throughout these regions. This means that, when we do find shared innovations, they are likely to date back to before the wider dispersion throughout Europe, specifically, prior to the separation from the western-central European core. The only isoglosses that are testimony to this very old coherence are the *j*-prothesis, the advance of *geljas* and renewal of *kones-*, and the generalisation of *av-*. This is the diffusion space that gives some reality to the notion of ‘Northern’ dialects – though it also includes the Northern Central dialects, and so it is not a definitive marker of a so-called ‘Northern branch’. Similarly, the cluster reduction in *maro* etc. spreads farther south, reaching the western Balkans. It too is a prominent feature of the northern European dialects, but not one that is limited to them.

Another core is situated in Transylvania and Wallachia. Many of the innovations that occur here are local, and help demarcate the Vlax group of dialects. But others are projected into neighbouring regions, most notably the Black Sea coast regions and in part northern Greece, as well as, to the north, the dialects of eastern Slovakia and southern Poland (Maps 3, 7). In relation to the diffusion spaces discussed above, Vlax is clearly defined the core of the southeastern diffusion zone.

From the more global perspective of diffusion zones throughout the Romani-speaking landscape, the divisions among the Balkan dialects appear in somewhat new light. The entire region is separated from the northern European dialects, and in some features also from Vlax, by the Great Divide. In the western Balkans, we find a zone that is connected, via the Adriatic coast region, to developments in central Europe, and shares a series of features found to the northwest. These include cluster reduction in *maro* etc., selection of *-h-*, selection of *-in-* in loan verbs (with central-eastern Europe), and in part the form of demonstratives. The Black Sea coast dialects participate in some of the developments that are typical of the southeastern core, spreading in and around Vlax (Maps 3, 7), but stand out, apart from regional developments, in the preservation of *-iz-* in loan verbs. Apart from those, the Balkans in general is a rather conservative area, its principal, global innovation being the partial analogy in the 2.PL perfective marker, *-en*.

The Northeastern group, or Polish-Baltic-Northrussian dialects, stand out through their participation in the more general changes that occur north of the Great Divide, while resisting a series of specific changes that are projected from the western-central diffusion core (Sinti). They constitute a diffusion space in their own right in the development of demonstratives and loan verb markers (Maps 7, 8), in addition to specific regional innovations.

The combination of the Sinti dialects of Germany and surrounding regions, with the Scandinavian dialects (represented by the present-day dialect of Finland), appears as perhaps the most coherent and distinct group (the Northwestern group). The dialects

pattern together not just in general features found north of the Great Divide, but also in a series of innovations (Maps 4, 5, 6, 7).

The Central dialects are split by the Great Divide, which separates the Northern from the Southern Central varieties. One of the few features typical of a Central diffusion space is the presence of demonstratives in *ada*, *aka*, which however often appear alongside other demonstratives. In other respects, the Central dialects are characterised by shifting affiliations with various diffusion spaces. They pattern with the Northwestern dialects in relation to *-h-* and *-al* (Maps 4, 5), and with the Northeastern group in relation to the loan verb marker *-in-* (Map 6), while the eastern dialects among the Northern Central are often receptive of innovations from the southeastern zone (Vlax) (Maps 3, 7).

Finally, in Britain, Iberia, Italy, and the narrow transition zone on the north-eastern Adriatic coast (southern Slovenia-Croatia), and partly Epiros, we find a series of peripheral dialects, which may or may not participate in various isoglosses. For a series of phenomena, they form relic areas, which prove useful in the relative dating of the sequence of innovations.

8. Conclusions

It is firstly noteworthy that linguistic structures in Romani are distributed within geographical patterns. This means that linguistic developments and changes were shared by neighbouring population groups, suggesting that contact between neighbouring population was responsible for their diffusion. This in turn shows that Romani populations did not migrate randomly, but, to the extent that they maintained itinerant occupations, they appear to have travelled within the containment of specific regions.

Next, the fact that structural variants are distributed geographically rules out that the changes that led to the variation occurred before the relevant groups settled in their present locations. Instead it suggests that groups first settled, then developed differences in speech. This is confirmed by the fact that we often find that the outer peripheries share

older forms (for example demonstratives), while the areas in the middle of the map show changed. This suggests that the relevant changes occurred after the period of Romani settlement in the western, northern, and central parts of Europe, which occurred from the late fifteenth century onwards. Since we have documentation of Romani from the eighteenth century onwards, which shows variation between the dialects that is very close to the variation found today, we can conclude that the major changes took place during the sixteenth and seventeenth centuries.

Third, the fact that we can attribute the changes to this particular period help us explain at least some of the patterns, specifically those depicted on Map 2 (the Great Divide). The fact that there innovations on both sides of the Great Divide, which fail to spread across the divide, suggests interrupted contacts between the Romani populations on either side of the divide during the crucial period. This might be seen in connection with the rivalry between the Austrian and Ottoman empires during the same period. Map 3 shows the Vlax area as a dynamic zone, which influences its immediate surroundings both to the north and south.

Fourth, we see in Map 9 that a series of developments overlap to a considerable extent. They all involve the simplification, then re-structuring of complex morphological paradigms. Due to the complex and layered nature of the process, it is not surprising that similar patterns are shared by groups of populations that formed rather coherent communities over a longer period of time: The most obvious examples are the Sinti population of Germany and neighbouring regions, the Northeastern group, and the British and Iberian Romani populations (respectively). These similarities account for the impression of a tighter relation among the dialects in the respective groups cited above as a recent classification grid: Vlax, Balkan, Central, Northwest and Northeast.

Finally, the implications of the diffusion model for dialect classification are straightforward: dialects are likely to share a feature if they are on the same side of the demarcation line which marks the extent of geographical diffusion of a particular innovation (structural change). The affinity among individual dialects is thus not genetic, but one

that consists of the adoption of a set of similar features; hence it is not absolute, but gradual.

References

- Ackerley, F.G. 1914. The Romani speech of Catalonia. *Journal of the Gypsy Lore Society*, new series, 8, 99–140
- 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.
- 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.
- Boretzky, Norbert. 1998. Areal and insular dialects and the case of Romani. *Grazer Linguistische Studien* 50: 1-27.
- 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 210-276.
- Boretzky, Norbert. 1999b. *Die Verwandtschaftsbeziehungen zwischen den Südbalkanischen Romani-Dialekten. Mit einem Kartenanhang*. Frankfurt am Main: Peter Lang.
- Boretzky, Norbert. 2000. South Balkan II as a Romani dialect branch: Bugurdži, Drindari, and Kalajdži. *Romani Studies*, fifth series, 10: 105-183.
- Elšík, Viktor. 2000. Dialect variation in Romani personal pronouns. In: Elšík, Viktor, and Matras, Yaron. eds. *Grammatical relations in Romani: The noun phrase*. Amsterdam: Benjamins.65-94.

- Gilliat-Smith, B.J. 1915. A report on the Gypsy tribes of North East Bulgaria. *JGLS*, new series, 9: 1-54, 65–109.
- Matras, Yaron. 1995. Verb evidentials and their discourse function in Vlach Romani narratives. In: Matras, Yaron. ed. *Romani in contact: the history and sociology of a language*. Amsterdam: Benjamins. 95-123.
- Matras, Yaron. 1999. The speech of the Polska Roma: Some highlighted features and their implications for Romani dialectology. *JGLS*, fifth series, 9(1): 1-28.
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Miklosich, Franz. 1872-1880. *Über die Mundarten und Wanderungen der Zigeuner Europas X-XII*. Wien: Karl Gerold's Sohn.
- Paspati, Alexandre G. 1870 [reprint 1973]. *Études sur les Tchinghianés ou Bohémians de l'Empire Ottoman*. Osnabrück: Biblio.