

Yaron Matras

*S/H* ALTERNATION IN ROMANI:  
AN HISTORICAL AND FUNCTIONAL INTERPRETATION<sup>1</sup>

1. INTRODUCTION

Three principal questions are addressed in this paper: 1) What is a useful way to classify Romani dialects? 2) How relevant is the areal spread of linguistic innovations in a language like Romani, which lacks territorial coherence? And 3) Based on contemporary dialect differences, can a case be made against the structural coherence of Romani in its very early stages; in other words, is there any evidence that present-day variation in Romani derives either from a mixture of Indo-Aryan "parent" languages, or from historical variation carried forward from a pre-European, Indo-Aryan Proto-Romani?<sup>2</sup> While these questions are linguistic in nature and concern the interpretation of a linguistic-structural state of affairs, their implications may range beyond linguistic issues, touching on the question of the origin and unity of Romani-speaking populations, and especially on the types of social and regional contacts among Romani populations within Europe that will have allowed or constrained the diffusion of linguistic innovations in particular regions.

My conclusions will be firstly that in considering the *s/h* variation across Romani dialects, we are dealing with several different structural developments that may or may not overlap. Some of these developments are likely to have their origin in underlying variation in earlier Indo-Aryan stages, though there is no evidence that current dialect boundaries within Romani reflect underlying dialects or language boundaries in India. Indeed, the most relevant developments show clear regional patterns of diffusion within present-day Romani in Europe. From this, one is tempted to conclude that the spread of innovations in Romani follows a similar pattern as in languages that are geographically coherent – "territorial" languages –, with Romani dialects forming a linguistic-geographical continuum. This in turn implies regular historical contacts between contiguous Romani populations across

---

1 Acknowledgements: An earlier version of this paper was presented at the Annual Meeting of the Gypsy Lore Society, 28-29 June 1999, in Florence. For helpful comments and discussions I am grateful to Peter Bakker, Donald Kenrick, Victor Friedman, Zoran Lapov, and Viktor Elšík.

2 Throughout I use the following terminology: "Proto-Romani" pertains to the pre-European stage for which some form of a more-or-less coherent Indo-Aryan ancestor of present Romani can already be assumed. "Early Romani" refers to the stage after the arrival in the Balkans or Asia Minor, and just before the spread of dialects beyond this area (usually to the north and west, but in some cases also eastwards); this is the stage during which the syntactic Balkanisation of Romani takes place (cf Matras 1994). "Common Romani" refers to general features shared by the majority of present-day Romani dialects.

Europe. Such evidence must be taken into account when discussing Romani populations in the context of nomadic and migratory groups, for it clearly shows the limits of nomadism and the overall regional fixation of individual communities. With respect to a methodology of dialect classification in Romani, then, the *s/h* alternation supports the view that isoglosses representing shared innovations (as well as, potentially, conservativisms) are likely to give a more accurate and differentiated picture of historical connections among present-day dialects and their forerunners than attempts at wholesale genetic groupings. Notwithstanding, more recent migrations, common among the Romani population as a whole, must of course be taken into account when mapping isoglosses. Whatever the present geographical position and the migratory history of a given Romani speech variety is, however, its features are best defined by an index of isogloss abstractions representing linguistic innovations that are potentially shared with the respective contiguous varieties from each and every historical stage in its development. A final conclusion pertains to the system-related characteristics of the *s/h* alternation. The observation outlined below is that the change from *s > h* in Romani is functionally motivated and hierarchically structured, and that it therefore follows a course that is, once set in motion, on the whole predictable. This finding is relevant for a more general theory of markedness and motivation for structural change in language.

## 2. WHAT IS THE *S/H* ALTERNATION IN ROMANI?

Let us begin by examining the problem. Figure 1 offers an insight into *s/h* alternation in Romani grammatical paradigms, featuring five selected dialects. The Welsh dialect is considered part of the Northern branch, though it shares many characteristics with the southern Balkan dialects. Sinti or German Romani is likewise part of the Northern branch of dialects, sharing some lexicon and morphological features with such dialects as Welsh, Finnish, or the Baltic-Northeastern group. Roman (spoken in the Austrian Burgenland and Slovenia) belongs to the Central group. Lovari is Transylvanian Vlax, while the dialect of the Montenegrinian Das (Orthodox Christians) represents the group of southern Yugoslav varieties of Vlax (akin to Gurbet and Džambazi) which appear to have emigrated into the area sometime in the 19<sup>th</sup> century at the latest. Though not a complete inventory, Figure 1 nevertheless provides a concise overview of forms and categories that are likely to be affected by the *s/h* alternation.

We find firstly stems for the copula, in various persons and in both tenses (present and past). Next we find the interrogative 'what' representing a class of interrogatives and determiners in *s/h* (see below). Both these categories, copula verbs and interrogatives, show *s/h* alternation in initial position. The long forms of the present verb conjugation show *s/h* in intervocalic position for the 2sg and 1pl; these forms denote either the present tense or

the future, or both (they are lacking in Montenegrinian Das, which has short or *a*-less forms for the present, and a particle *ka* for the future). Likewise, in the case system we find the instrumental with intervocalic *s/h*. The forms that follow represent *s/h* and *s/∅* alternation in final position. We have the m.sg accusative case; the tense suffix which I call the "remoteness marker" and which forms the imperfect when added to the present and the pluperfect/contrafactual when added to the past tense; the concord marker for the 3sg past tense of the verb; the so-called athematic ending for m.sg nouns in *-os* (as well as *-is*) – a Greek-derived nominal ending that is added to European loanwords (see Bakker 1997); and finally the final consonant in lexical words, represented here by the word for 'day'.

Figure 1: *s/h* alternation in selected Romani dialects

	Welsh	Sinti	Roman	Lovari	Montenegr. Das
Cop 1sg pres	<i>som</i>	<i>hom</i>	<i>som</i>	<i>som</i>	<i>sem</i>
Cop 2sg pres	<i>san</i>	<i>hal</i>	<i>sal</i>	<i>san</i>	<i>san</i>
Cop 3 pres	<i>si</i>	<i>hi</i>	<i>hi</i>	<i>si</i>	<i>hi</i>
Cop 3 past	<i>sas</i>	<i>his</i>	<i>sina</i>	<i>sas</i>	<i>seha</i>
Interr 'what'	<i>so</i>	<i>ho</i>	<i>so</i>	<i>so</i>	<i>so</i>
Verb 2sg (pres long)	<i>-esa</i>	<i>-eha</i>	<i>-eha</i>	<i>-esa</i>	-
Verb 1pl (pres long)	<i>-asa</i>	<i>-aha</i>	<i>-aha</i>	<i>-asa</i>	-
Instr m.sg, f.sg	<i>-esa, -asa</i>	<i>-eha, -aha</i>	<i>-eha, -aha</i>	<i>-esa, -asa</i>	<i>-eha, -aha</i>
Acc m.sg	<i>-es</i>	<i>-es / -e</i>	<i>-es</i>	<i>-es</i>	<i>-eh</i>
Imperf/pluperf=rem	<i>-as</i>	<i>-s</i>	<i>-ahi</i>	<i>-as</i>	<i>-a</i>
Verb 3sg past	<i>-as</i>	<i>-as</i>	<i>-a</i>	<i>-as</i>	<i>-a</i>
Athem m.sg	<i>-os</i>	<i>-o</i>	<i>-o</i>	<i>-o</i>	<i>-o</i>
'day'	<i>dives</i>	<i>dives</i>	<i>di</i>	<i>džes</i>	<i>diveh</i>

Figure 1 already demonstrates the difficulties in integrating the *s/h* alternation into an overall dialect classification scheme. Although the Welsh dialect seems to have *s* consistently, there appears otherwise to be no clear-cut distinction between *s*-dialects and *h*-dialects. Moreover, we see that *s/h* alternation may affect different forms in different dialects. Most conspicuously we find that Sinti is consistent in showing *h* in initial and medial positions but has *s* in final position, while Das has only some *h* forms in those positions, but is consistent in dropping *s* in final position. This suggests that we are in fact dealing with (at least) two separate phenomena, which are analogous to Ferguson's (1990) distinction between a "Greek type" and a "Spanish type" of *s>h>∅* phonological development, where *s* is firstly aspirated, then eventually dropped. In the "Greek type", *s* is reduced in first instance in intervocalic position, the change then proceeds to word-initial position (especially at first when the preceding word ends in a vowel), then to pre-consonantal positions, and only then, if at all, to word-final position. This reminds us at first glance of what we find in

Sinti, where intervocalic and initial positions are affected, but final (and potentially preconsonantal) positions only marginally so (in the accusative m.sg, but not elsewhere). In Ferguson's "Spanish type", the development starts with syllable-final and preconsonantal positions, then proceeds to word-final positions, and will affect last if at all word-initial positions. This is similar to the case of Das, where *s* in all final and intervocalic positions is reduced, but only marginally do we find *s* reduction in word-initial position. Lovari seems hardly affected by this process, with *s* reduction only in athematic noun endings. The process is further developed in Roman, though not all final positions are affected and we have *s*-retention in the accusative m.sg. Can *s/h* alternation in Romani then be interpreted as two distinct processes of phonological *s* reduction, the first, of the Greek type, spreading in western-central Europe as exemplified by Sinti, and the second, of the Spanish type, spreading from southeastern into central-eastern Europe, as exemplified by the Das-Roman continuum, and marginally affecting a contiguous variety such as Lovari?

Disregarding for the moment the strength of the geographical argument, for which Figure 1 admittedly only provides a very narrow set of data, let us first examine the applicability of the phonological argument. We begin with the postulation of historical *s* and so of a loss of *s* where it is missing in cognate expressions. This seems reasonable for final positions, both on grounds of universal phonetic processes, and when comparing the patterns of variation exemplified in Figure 1; thus, reduction of final *s* appears to be a gradual process. Furthermore, we can ascertain underlying etymological *s* in items like *dives* 'day' as well as for the accusative m.sg ending (OIA genitive in *-s*) and the athematic ending (Greek *-os*). The principal difficulty lies with the change in medial and initial positions. Here, we are not dealing with a regular sound change, but with one that affects strictly grammatical paradigms – inflectional endings and function words. Thus we do not find *s* reduction in a word like *sap* 'snake', or *resela* 'to arrive'. There appear to be only two cases of lexical cognates with *s/h* alternation, both are all but straightforward cases of regular and predictable phonological development. The first concerns the Common Romani stem *šun-* 'to hear'. Its cognate in some dialects of Sinti is *hun-*. Now, OIA *ś* (*śrn-* 'to hear') evolves regularly to *š* in Proto-Romani; we have in present-day Romani no *š/h* alternation, so that for Sinti either an underlying borrowing from another Indic language with OIA *ś* > *s* (possibly Domari) might be postulated, with a subsequent development *s* > *h* analogical to that observed above for grammatical paradigms; or else an even more exceptional analogical shift of *š* directly to *h* must be assumed. In either case, we are dealing with an exception that does not justify an interpretation as the beginning of a spread of the *s* > *h* shift to lexical items, despite the fact that it is in Sinti that we find the most consistent "Greek type" *s* > *h* development. The second case is the alternation of *taisa* 'tomorrow, yesterday' and *tehara, taharin* 'tomorrow, in the morning'. This item of unclarified etymology is explained by Boretzky/Igla (1994) as

a borrowing of Greek *taxiá* 'morning', calquing an underlying Indic expression *kal-* which signifies both 'tomorrow' and 'yesterday'. The sound alternation in Romani might be explained in terms of the rarity of medial, palatal *x* in Greek expressions borrowed into Early Romani. Moreover, the split between predominantly Vlax *teh-/tah-* and non-Vlax *taisa* in both form and semantic meaning would suggest two separate borrowing processes rather than a borrowing into Early Romani with subsequent phonological and semantic differentiation, which in turn would imply a rather early split of the Vlax dialects from Early Romani.

To return to the main line of our discussion, we can establish that 1) there is evidence in favour of a phonological development of the "Spanish type" affecting *s* in final and perhaps also medial position in some dialects of Romani; I shall call this development "Process B", as its roots appear to have emerged at a somewhat later stage (see discussion below), 2) there is a separate development involving *s/h* alternation in initial and medial position which affects strictly grammatical paradigms, but not lexical items, and which therefore cannot be taken to reflect a phonological change from underlying *s* to *h* without further evidence. In the following section I explore the historical roots of this latter development, to which I shall refer as "Process A".

### 3. HISTORICAL EVIDENCE FOR *S/H* ALTERNATION IN INDO-ARYAN

How likely is it that *s/h* alternation in Romani might reflect earlier dialect divisions within Indo-Aryan? In fact we do find variation in Indo-Aryan which could be compared to the Romani situation. Beames (1872 Vol I: 258) discusses the difficulties in tracing the geographical origin of the underlying shift from *s* to *h* which, due to the impact of migrations in medieval times has become a feature of so many different varieties and so a characteristic of modern Indo-Aryan as a whole, rather than of geographically confined isoglosses. In particular, Beames mentions in this connection the Jats and Rajputs, both migrating mixed populations which have often been associated with the ancestors of the Roma. The striking feature of *s/h* alternations in modern Indo-Aryan is that they are confined to particular morphological paradigms, rather than to particular regions, and in this they differ from more general phonological developments from *s > h* as found for example in modern Singhalese (cf. Masica 1991: 187, 206). Turner (1927 [1975]) argues that no satisfactory explanation can be given for the historical development of forms in *-h-* for the underlying MIA future suffix in *-isya* other than the special treatment of *s* in a terminational element. A further example is the treatment of the MIA 2sg indicative present in *-asi* which changes via *-ahi* to the vocalic endings in NIA languages such as Gujarati and Hindi, which have otherwise retained *s* in the body of a word. Turner's conclusion is that consonants in grammatical

terminations are liable to a development that is different from that of the same sounds in the body of the word. He relates this to the functional decline of inflectional terminations, resulting in a weakening pronunciation and the reduction or loss of consonants as a consequence of reduced muscular effort.

As a phonological phenomenon, fricative weakening resulting in  $s > h$  starts operating in the transitional period between OIA and MIA, appearing before and after voiceless stops (OIA *akṣi* 'eye' > MIA *akkhi*), before nasals (OIA *asme* 'we' > MIA *amhe*), and marginally also in intervocalic position (cf. Bubeník 1996: 46–51). It gains ground, in all likelihood initially in the spoken language, especially in grammatical paradigms, appearing already in literary dialects used by Jaina authors for poetic works between the 5<sup>th</sup>-12<sup>th</sup> centuries, the varieties termed *Apabhraṃśa*, which represent the latest and, in a sense most colloquially-influenced forms of Middle Indo-Aryan (cf. Bubeník 1998: 50–87). Compared to the earlier Prākṛits, which have forms in *s*, the late Prākṛits or *Apabhraṃśa* show forms in *h* for a number of grammatical formants, as seen in Figure 2:

Figure 2:  $s > h$  in the emergence of late MIA (*Apabhraṃśa*) grammatical paradigms

locative singular	<i>-a-smin</i> > <i>-a-him</i> > <i>-hi</i>
1pl pronoun	<i>asme</i> > <i>amhe</i>
cop 1sg pres	<i>asmi</i> > <i>(a)mhi</i>
genitive/dative singular	<i>-assa</i> > <i>-aha</i>
Verb 2sg pres	<i>-asi</i> > <i>-ahi</i>
future marker	<i>-isya</i> > <i>-ha</i>

The shift is not uniform, however. Already in pre-*Apabhraṃśa* texts, namely in *Māgadhī*, there is evidence of a locative singular *-a-him* > *-hi* from OIA *-a-smin*, as well as of a future marker in *-ha-* from OIA *-sya-* (see Tagare 1948 [1987]: 130). It seems that analogies within the grammatical paradigms are responsible for the gradual diffusion of  $s > h$ . Bubeník (1996: 104–110) postulates a spread of  $s > h$  in verbal paradigms beginning with the 2sg future *-issasi* > *-ihisi*, which emerges through a dissimilatory process eliminating two intervocalic *s*'s, with *-h-* then being re-interpreted as the future marker and generalised throughout the person paradigm. The process later affects the 2sg in *-asi* > *-ahi*. Curiously, this latter process is later reversed with the revival of Brahmanism in the 12<sup>th</sup> century and the re-introduction of Sanskrit (OIA) forms; there is however a geographical distribution here, with Western *Apabhraṃśa* *-ahi* leading to reduced 2sg forms in the NIA languages of this region, e. g. Gujarati and Hindi, while in Eastern *Apabhraṃśa* *-asi* prevailed, still represented in eastern NIA languages such as Bengali and Maithili (cf. Tagare 1948 [1987]: 287–288). The effects of this reversal in the written language are illustrated in Figure 3:

Figure 3: Variation within MIA

Incidence of Apabhraṃśa *-ahi/-ēhi* vs Prākṛit *-asi/-ēsi* (2<sup>nd</sup> pers sg) in late Apabhraṃśa works (reproduced from Alsdorf in Tagare 1948 [1987]: 288; Bubeník 1998: 62)

	<i>-ahi/-ēhi</i>	<i>-asi/-ēsi</i>
Dhaṇavāla (10 <sup>th</sup> cent)	48	2
Haribhadra (1159)	1	regular
Somaprabha (1195)	6	12

A further type of *s/h* alternation in NIA which might be compared with that found in Romani copula paradigms results from lexical selection in the stem used for the copula auxiliary. Among the most common forms in NIA are derivations of OIA *ačč->čh-*, *sth->th-*, *as->s-*, and *bhū->h-*. The distribution of these various forms may be regional as well as paradigmatic, creating suppletive paradigms (e. g. Hindi present tense in *h-*, past tense in *th-*). Might it be possible that such variation was carried over into Proto-Romani, later to become regularised in different ways among different Romani dialects? Woolner (1924: 182) argues against the continuation from Indic *h-* forms to the Romani copula in *h-* claiming that Romani drops Indic initial *h-*. He suggests instead that *hi* etc. emerged from an enclitic *si* which underwent reduction, and was then copied by the independent copula. Boretzky (1995: 17) contradicts Woolner, pointing out that initial *h-* in Romani is in fact occasionally preserved, and besides that the change *s>h* in Indic is not a strictly phonological one either, but is confined to grammatical paradigms, hence, implicitly at least, exemption of the copula from phonological developments might be expected. Boretzky's alternative scenario regards the *s/h* variation in the copula as a trace of two separate paradigms, an accentuated one in *s-* and a de-accentuated one in *h-*, having its origins in variation within MIA (*asmi>amhi* 'I am'). The two paradigms are then combined into one, with dialect-specific outcomes. This means that forms in *h-* and in *s-* can be taken to have coexisted, until dialect-specific levelling led to a choice of just one variant per function (person, tense), resulting in either *s-*, *h-*, or else in mixed paradigms.

The elimination of *h-* forms where dialects generally opt for the *s-* paradigm is however not arbitrary, but follows a path that reflects functional markedness. Thus Boretzky (1995: 25) observes that some dialects have *s-* in the 1<sup>st</sup> and 2<sup>nd</sup> persons of the present tense, but *h-* in the 3<sup>rd</sup> person. In the past tense, dialects will tend to have a uniform paradigm, but where a mixture is encountered, it is again the 3<sup>rd</sup> person that may show *h-*. (I shall return to this issue of hierarchical spread and markedness below). In support of his analysis Boretzky mentions the presence of de-accentuated enclitic copula forms in *h>∅* even in those Vlax dialects which otherwise show copulas in *s* (*mišto-i* 'it is good'), an argument which in itself is not at all entirely incompatible with Woolner's suggestion. But he also points out the

retention of double forms of the copula – *sin-* alongside *hin-* – in (Kosovo) Arli, on the whole a rather conservative dialect. Further evidence not considered by Boretzky but which might support his explanation is given by Miklosich (1877: 59), who cites the use of copula forms in *h* in modal expressions in dialects where the general copula paradigm has *s*: thus in Finnish Romani, *hom te* serves as a future marker, and in Caló (Spanish-based Para-Romani), where forms in *sin-* appear, *chomte* has likewise survived in the modal *chonte penar* 'one must say'. The overall scenario suggested by Boretzky might thus be elaborated on as follows: Proto-Romani inherits MIA variation in the pronunciation of the copula paradigm, derived from OIA *as-*. The more recent forms in *h* are de-accentuated, and survive in some dialects only in enclitic position or in frozen modal expressions. In other dialects, namely in Arli, the two paradigms become functionally equivalent and continue to coexist. In others, the forms are re-arranged into a single paradigm, eliminating either the entire set in *s*, or that in *h*, or else combining the two without doubling the forms. The hierarchical structuring of the combined paradigm – with forms that are high in markedness tending to show *s* rather than *h* – is a leftover from the earlier functional differentiation, where *s*-forms were accentuated and *h*-forms de-accentuated.

While the notion of paradigm competition entailed by this scenario seems acceptable, there are however some additional considerations to be made in connection with the *s/h* tension in Early Romani copula paradigms and in particular the alternation *within* one single MIA lexical paradigm in *<as-*. The first is a structural consideration. The aspiration of *s* – or fricative reduction – in the MIA copula paradigm in *as-* is more clearly a phonological process than the *s/h* alternations in other grammatical paradigms illustrated in Figures 2 and 3. It is often associated with metathesis (*asmi* > *amhi* 'I am', *asti* > *athi* 'he is'), though this is not always the case. A simplification of *asti* > *athi* > *hi* for the 3sg is indeed conceivable. However, for the 1sg *asmi* > *amhi* we have a parallel case in the form of the 1pl pronoun *asme* > *amhe*, which in Romani (as well as in other NIA languages) consistently gives *ame*. It is thus more difficult to imagine a continuation of a de-accentuated MIA 1sg copula *amhi* into Romani *hom*. Furthermore, where we do find *s/h* alternations in copula paradigms elsewhere in NIA it is mostly the result of suppletion, as described above. Domari, for examples, has retained a conservative form of OIA *as-* in *ašti* 'there is', which in Syrian Domari also figures in the auxiliary enclitic copula *štum* 'I am' etc., while Jerusalem Domari generally employs a form in *hō-* for the enclitic copula as well as for the verb 'to become', clearly a derivation of OIA *bhū-* 'to become'. In addition, there are traces of an existential clitic in *hi*, which might be of a different source, namely late MIA *athi* > *hi*. Noteworthy is also the fact that while (central) Syrian and Palestinian Domari are VO languages (like coterritorial modern Arabic), the copula (which lacks a structural counterpart in Arabic) has remained strictly enclitic. In Romani the copula paradigm shows a unique alternation of the vowel in the 1sg ending between *-om* (non-Vlax) and *-em* > *-im* (Vlax), while



in the 3<sup>rd</sup> person the form is consistently *-i* (*si* or *hi*). We cannot exclude therefore that all three sources may have contributed to the Romani copula system: A stem in *s*- < OIA *-as*, a stem in *ho/he* < OIA *bhū-*, and a reduced 3<sup>rd</sup> person enclitic copula in *hi* < MIA *athi*.

Turning now to a summary of the historical evidence considered so far, let us review the possibilities for a takeover of late MIA *s/h* variation in grammatical paradigms into Proto-Romani (Figure 4):

Figure 4: Possible continuation of historical *s>h* in Romani

Category:	MIA (Apabhramṣa) form:	Cognate in Romani:
locative singular	<i>-a-smin</i> > <i>-a-him</i> > <i>-hi</i>	NONE
future marker	<i>-isya</i> > <i>-ha</i>	NONE
1pl pronoun	<i>asme</i> > <i>amhe</i>	<i>ame</i>
genitive/dative singular	<i>-assa</i> > <i>-aha</i>	<i>-es</i> ( <i>-eh</i> )
cop 1sg pres	<i>asmi</i> > ( <i>a</i> ) <i>mhi</i> // <i>as-</i> , <i>bhū</i>	<i>som</i> , <i>hom</i>
cop 3 sg pres	<i>asti</i> > <i>athi</i>	<i>si</i> , <i>hi</i>
Verb 2sg pres	<i>-asi</i> > <i>-ahi</i>	<i>-esa</i> , <i>-eha</i>

Of the grammatical endings that show *s/h* alternation in MIA, neither the locative marker nor the future marker have cognates in Romani (the Romani future in *-a* being the long form of the present, specialised to mark the future in some dialects that employ short forms as present indicative rather than subjunctive). The 1pl pronoun appears in Romani consistently in a reduced form that bears no trace of an *s/h* variation or *s>h* shift. A candidate for a carry-over of *s/h* alternation might be the MIA genitive/dative m.sg in *-assa>-aha*. This form gives rise to the general oblique in the m.sg, its consonantal form in *s* being retained only in peripheral NIA languages such as Romani, Kashmiri, and Domari, while elsewhere full reduction to a vowel ending has taken place. We do not need to find a MIA correlate for the *s/h* alternation in this form in Romani, however, since this form is more likely to be affected by the strictly phonological development of final *s* reduction ("Spanish type"), though some occurrences are found in Sinti as well.

In the copula, the evidence offers several scenarios and possibly a combination of processes that could be responsible for present-day *s/h* alternation in Romani, including the co-existence of *s*- and *h*-paradigms in accentuated and de-accentuated functions respectively, the co-existence of two competing or suppletive lexical sources, the separate survival of a de-accentuated enclitic copula, or, perhaps the most likely scenario, a mutual contamination of co-existing copula forms. The final candidate for a carry-over of the MIA alternation is the 2sg present conjugation ending, represented in Romani by the long conjugation where *s/h* appears in intervocalic position.

This leaves several instances of *s/h* alternation in Romani unaccounted for simply in terms of possible continuation of late MIA variation. Firstly, there is the Romani instrumental case. This is a Romani innovation, paralleling the grammaticalisation of postposed *samau/sahu* in other NIA languages but leading in Romani (as in Domari) to the formation of an agglutinating instrumental suffix. Clearly, then, we must assume a Proto-Romani form in *s*. We might explain the emergence of *h*-forms of the instrumental as an analogy to the verb conjugation, where we find identical sequences, namely in *-esa*, for both 2sg (long) present tense concord forms and m.sg instrumental; the analogy may then have spread to include the f.sg instrumental in *-asa>aha*, as well as the 2sg pronominal form *tusa>tuha* 'with you', while in postconsonantal position the instrumental retains *s* throughout (*lensa* 'with them' etc.). A similar analogous development might explain a further feature of the Romani alternation that has no apparent cognate in MIA, namely the plural concord in the present conjugation *-asa>-aha*. Here again, an analogy to the 2sg verb form might be assumed, perhaps in conjunction with the spread of *s>h* to the instrumental affecting first the m.sg and then the f.sg in *-asa>-aha*, and with the general patterns being toward consistent reduction of *s* in intervocalic position in grammatical paradigms. No such explanation offers itself at first glance for the interrogatives and determiners, nor for the reduction of *s* in final position in Romani. In both cases, unlike the other instances of *s/h* alternation in present-day Romani, we have no evidence for variation or causes for variation in Proto-Romani. The next section will deal with the dynamics of the *s>h* development within Romani itself.

#### 4. THE HIERARCHICAL NATURE OF *S>H* IN GRAMMATICAL PARADIGMS ("PROCESS A")

Figure 1 already illustrated the uneven distribution of *s>h* development across various grammatical forms. Let us consider the categories one by one. Figure 5 summarises the principal features obtained from a comparative investigation of ca. 50 dialect descriptions (see Appendix):

Figure 5: Copula forms (summarised)

	3 pres	3 past	1&2sg pres	1&2sg past
Sinti group	<i>hi</i>	<i>his</i>	<i>hom, hal</i>	<i>homs, hals</i>
Slovak & S. Polish	<i>hi</i>	<i>has</i>	<i>som, sal</i>	<i>somas, salas</i>
E. Slovak, Finnish	<i>hin</i>	<i>sas</i>	<i>som, sal</i>	<i>somas, salas</i>
Montenegr. Das	<i>hi</i>	<i>seha</i>	<i>sem, san</i>	<i>sema, sana</i>
S. Central	<i>hi</i>	<i>si(n)a</i>	<i>som, sal</i>	<i>somahi, salahi</i>
Arli	<i>si/hi</i>	<i>sine/hine</i>	<i>s/hinum, s/hinan</i>	–
All other	<i>(i)si</i>	<i>sas, sis, sine</i>	<i>s(in)som, s(in)an</i>	<i>s(in)somas, s(in)anas</i>

Taking into consideration only the copula stem, and disregarding for the moment tense affixes, we find that mixed paradigms will always have a 3<sup>rd</sup> present form in *h*. This pertains of course to (Kosovo) Arli, which has a symmetrical copula paradigm throughout, and to the Sinti group, which has forms in *h* throughout. But more significantly it holds for the other groups of dialects with partial *s>h* developments: Thus in Eastern Slovak, Finnish, Das, and the Southern Central dialects there is only one copula form in *h*, namely that of the 3<sup>rd</sup> present. In Slovak and Southern Polish varieties, *h* forms are likewise restricted to the 3<sup>rd</sup> person but include the past as well as the present forms. Thus a form in *h* in the 3<sup>rd</sup> past necessarily entails *h* in the 3<sup>rd</sup> present, while *h* in the 1<sup>st</sup> and 2<sup>nd</sup> persons necessarily implies *h* in the 3<sup>rd</sup> person in both past and present:

- (1) cop 3 pres > cop 3 past > cop 1&2 (sg&pl) pres&past

This clearly favours a spread of *h* beginning in the least marked form, that of the 3<sup>rd</sup> person, which strengthens the argument in favour of a de-accentuated (unmarked) paradigm in *h* competing with an accentuated (marked) set in *s* (cf. Boretzky 1995). Let us now proceed to examine correlations with other classes of items. Figure 6 shows the results of the sample for the copula, represented here by the form most likely to have *h* namely that of the 3<sup>rd</sup> person present, and the instrumental:

Figure 6: Copula and instrumental

	cop 3 pres	instr
Sinti, Central, Das, Arli	h	h
Baukols & Lalere	h	s
All other	s	s

Only two dialects are found to be inconsistent in their treatment of *s/h* for the two forms. Both have what could be defined as a transitional status, as they are isolated from the branches to which they belong (Baukols is derived from Sinti and spoken in Italy, Lalere is a Vlach-type dialect of Bohemia under heavy influence from Sinti). On the whole, the co-occurrence of *h* forms for both classes – copula and instrumental – and so some kind of connection can be postulated, while the possibility of a slight prominence of the copula must be admitted, either in the internal development or as a result of cross-dialect interference:

- (2) cop 3 pres > instr (though usually co-occurrence)

Figure 7 expands the comparison of various classes to include the other categories that are potential candidates for the *s/h* alternation in initial and intervocalic position, namely the long verb conjugation and the interrogative:

Figure 7: Co-occurrence of various forms

	cop 3 pres	instr	verb 2sg/1pl	cop 3 past	interr	cop1&2
Sinti	h	h	h	h	h	h
Central, Finnish	h	h	h	h/s	s/(h)	s
Montegr. Das	h	h	-	s	s	s
Macedon. Arli	h	h	h	s	s	s
Other	s	s	s	s	s	s

If one disregards the two aberrant dialects Baukols and Lalere, a generalisation can be made with respect to the clustering of *h*-forms, with an interdependency of the following classes: the 3<sup>rd</sup> person present copula, the instrumental, and the long verb conjugation. This is the case in a variety of dialects in different geographical locations, notably central Europe, the northern Baltics, and the southern Balkans. The more frequent presence in the sample of *h* in the 3<sup>rd</sup> person past-tense copula, in relation to the marginal presence of the interrogative in *h* in the Central dialects, justifies placing the former – cop3past – higher on the hierarchy for *s* > *h* than the latter – the interrogative, while 1<sup>st</sup> and 2<sup>nd</sup> person copula remains in lowest position:

- (3) cop 3 pres, instr, long verb conj > cop 3 past > interr > cop 1&2

Finally, a slight modification of the hierarchy may be added, pertaining to the internal hierarchisation among interrogatives and determiners, presented in Figure 8:

Figure 8: *s* > *h* in interrogatives and determiners

	'how'	'which'	'what'	'what for'	'every'	'all'
Lálere	<i>sar</i>	<i>havo</i>				
Slovak	<i>sar/har</i>	<i>havo</i>	<i>so</i>	<i>soške</i>	<i>sako</i>	
Moravian		<i>havo</i>	<i>so</i>		<i>sako</i>	
Radnitz	<i>har</i>	<i>havo</i>	<i>so</i>	<i>soške</i>	<i>sako</i>	
Hameln Sinti	<i>har</i>	<i>savo</i>	<i>hoi</i>	<i>hoske</i>		<i>ha-lauter</i>
Rhine Sinti	<i>har</i>	<i>savo</i>	<i>hoi</i>			
Marburg Bible Sinti	<i>har</i>	<i>sawo</i>	<i>hoi</i>	<i>hoske</i>	<i>hako</i>	<i>ha-lauter</i>
Bohemian Sinti	<i>har</i>	<i>havo</i>	<i>ho</i>			
Sinti	<i>har</i>	<i>havo</i>	<i>ho</i>	<i>hoske</i>	<i>hako</i>	<i>ha-lauter</i>
Rüdiger's Sinti	<i>her</i>	<i>havo</i>	<i>ho</i>		<i>hako</i>	<i>ha-</i>
Auvergne Manush	<i>har</i>		<i>ho</i>	<i>hoske</i>	<i>ako</i>	
Lombard Sinti	<i>ar</i>		<i>o, so</i>			
Venetian Sinti	<i>har</i>	<i>havo</i>	<i>ho</i>		<i>hako</i>	

Clearly,  $s > h$  in interrogatives and determiners is a feature of the Sinti group, influencing marginally adjoining northern-Central dialects, where the incomplete data set suggests that 'how' and 'which', but not the other forms, are susceptible to  $s > h$ . This contradicts somewhat the finding for the second group of central-northern dialects of Sinti, where all forms but 'which' have  $h$ . The final grouping shows the predominance of  $h$  (or  $\emptyset$ ), with  $so$  alongside  $o$  in Lombard Sinti. More conclusive despite the gaps in the data set is the relation between interrogatives and determiners. We have two determiners that participate in  $s > h$ : the first,  $s(v)ako$  'each', is a Slavic borrowing into Early Romani and so it helps us date the  $s > h$  development with certainty to the period following the dispersal of Romani-speaking populations from the Balkans westwards, from the 14<sup>th</sup> century onwards. The second,  $sa$  'all', is usually reinforced in Sinti by the German loan *lauter*, a classificatory item indicating the entirety of a class of objects. As far as can be maintained on the basis of the sample, the two determiners appear to be linked. More importantly, a determiner will not undergo  $s > h$  unless also an interrogative, and more likely the majority of interrogatives in historical  $s$  have also undergone the same development. This allows us to insert the following modification to the hierarchy:

- (4) cop 3 pres, instr, long verb conj > cop 3 past > interr > determ > cop 1&2

## 5. AN INTERIM SUMMARY

There are clear cross-dialectal regularities in the spread of  $s > h$  throughout the grammatical system ("Process A"); they enable us to predict the extent of  $s > h$  in a given dialect, as well as to reconstruct the general course which the underlying historical development has taken. Only some of these regularities can be attributed to phonetic factors or the regularity of sound change in similar environments: notably the correlation between long forms of the verb and the instrumental. Here one can assume that the 2sg form  $-esa > -eha$  triggers an analogical development in the instrumental m.sg  $-esa > -eha$ . Both forms then spread within their respective paradigms, in the verb conjugation to encompass the 1pl  $-asa > -aha$  and in the nominal paradigm to the instrumental f.sg  $-asa > -aha$ . In other items, markedness and frequency appear to play the decisive role, with lower markedness and higher frequency correlating positively with the likelihood to be affected by the  $s > h$  development. This explains the prominence of the 3<sup>rd</sup> person over 1<sup>st</sup> and 2<sup>nd</sup> persons in the copula, the present over the past, the copula over interrogatives, and interrogatives over determiners.

We have already defined Process A as being confined to grammatical paradigms and so distinct from the purely phonological reduction of word-final  $s$ . We are now in a position to identify at least three distinct sub-components within Process A: 1) analogous phonetic

reduction of intervocalic *s* in grammatical endings (verb present conjugation and nominal inflection), 2) competition between two coexisting sets for the copula (being the historical outcome either of phonological reduction in de-accentuated forms, or of different lexical selection, or both), 3) an analogous spread of *s>h* substitution to other function words in *s*.

The overall development labelled here as Process A is in need for both a trigger and a motor, which drives the process along a predictable and consistent course until its maximal realisation, in some of the Sinti dialects. As a trigger we may postulate the presence of inherited variation, most likely in fact of several patterns of inherited variation – one resulting from a reduction of *s* in inherited morphological paradigms, most likely in the verb conjugation, another consisting of alternating reduced forms and *s*-forms in the copula, and another reflecting parallel grammaticalisations of different lexical items as part of the copula paradigm. The motor driving the development toward uniformity which favours *s* is the need for regularity and consistency within the affected paradigms, and beyond those, among classes of words that are interpreted as potentially analogous. How does such "potential analogy" emerge? Indeed, it is the cross-paradigmatic diffusion of *s>h* which constitutes the most striking feature of the *s/h* alternation in Romani. Two factors are needed for this diffusion to go ahead. The first is of course the presence of a candidate to undergo an analogous sound substitution – namely an *s*. Although it is not the articulatory process that is the driving force behind *s>h* in these domains, it is important that candidates for *s>h* substitution appear in environments that are not resistant to a normal course of phonetic change; this exempts postconsonantal *s*'s in the instrumental paradigm (*mansa* 'with me', *lensa* 'with them', etc.). The second factor is a perception of elements as belonging to the grammatical apparatus of language, rather than to its lexicon. It is this mental categorisation that allows speakers to draw analogies between such loosely-related forms as the copula, noun and verb inflection, interrogatives and determiners, while skipping lexical material.

We have so far attempted to explain the less obvious process of *s>h* diffusion in Romani dialects – Process A, which bears some superficial resemblance to Ferguson's (1990) "Greek type" though it is confined to grammatical markers and so it does not constitute a sound change in the strict sense. The following section is devoted to what might be seen as the more straightforward pattern – the reduction and loss of *s* in word-final position or "Process B".

## 6. REDUCTION OF WORD-FINAL *s* ("PROCESS B")

Above we tentatively identified a second type of *s>h* development in Romani involving the reduction of *s* in word-final position, partly affecting preconsonantal *s* as well, and so resembling Ferguson's (1990) "Spanish type" sound change. In sharp contrast with the development dealt with in detail in the previous sections, *s*-reduction in word-final position in

Romani is characterised by the fact that it is a general development, affecting lexical items as well as grammatical endings. Furthermore, it can be safely assumed to be a rather late development, which allows us to postulate an historical *s* as the point of departure. This being said, it is nevertheless necessary to examine the precise extent of the process. Figure 9 provides an overview of some of the more conspicuous categories that are affected. Some of these were already mentioned in the introductory remarks to the phenomenon (section 1). They include the athematic nominal ending of the m.sg in *-os*, the concord marker for the 3sg in past tense of lexical verbs in *-as*, the accusative ending of the m.sg in *-es*, and word-final *s* in lexical items, exemplified here by the word for 'day'. Added to this list are in Figure 9 the (accusative) form of the reflexive pronoun in *pes*, the adverbial derivation markers in *-es*, and the short present conjugation form of the verb (subjunctive or present) for the 2sg in *-es* and 1pl in *-as*. Missing, though potentially affected, are the remoteness marker and the general oblique m.sg form, etymologically identical with the accusative but appearing in preconsonantal position. These will be dealt with separately.

Figure 9: *s* in word-final position

	athem m.sg	3sg past	refl	acc m.sg	adv	'day'	short verb 2sg& 1pl
Montenegrean Das	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>	<i>-eh</i>	<i>d'iveh</i>	∅
Prishtina Gurbet	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>		<i>džive</i>	∅
Scutari Vlach	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>		<i>givé</i>	∅
Kosovo Arli	<i>-o</i>	<i>-ja</i>	<i>pe(s)</i>	<i>-é</i>		<i>avdive</i>	∅
Macedonian Arli	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>	<i>-e</i>	<i>dive</i>	∅
Krajina Gurbet	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>	<i>-e</i>	<i>d'ive</i>	<i>s</i>
Roman	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>		<i>di</i>	<i>s</i>
Southern Central	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-e</i>	<i>-e</i>	<i>dive</i>	<i>s</i>
Romungri	<i>-o</i>	<i>-a</i>		<i>-e</i>		<i>dive</i>	<i>s</i>
Istrian Sinti		<i>-a</i>		<i>-e</i>		<i>dive</i>	
Kalderaš	<i>-o</i>	<i>-a</i>	<i>pe</i>	<i>-es</i>	<i>-es</i>	<i>gjes</i>	<i>s</i>
North Russian	<i>-o</i>	<i>-a</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>dēves</i>	<i>s</i>
Russian (North)		<i>-a</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>dyves</i>	<i>s</i>
Polska Roma	<i>-o</i>	<i>-a</i>	<i>pes</i>	<i>-s</i>	<i>-es</i>	<i>dives</i>	<i>s</i>
Hameln Sinti	<i>-o</i>	<i>-as</i>	<i>pes</i>	<i>-e</i>	<i>-es</i>	<i>dives</i>	<i>s</i>
Marburg Bible Sinti	<i>-o</i>	<i>-as</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>diwes</i>	<i>s</i>
Rüdiger's Sinti	<i>-o(s)</i>	<i>-as</i>		<i>-es</i>	<i>-es</i>	<i>dives</i>	
Lovari	<i>-o</i>	<i>-as</i>	<i>pe</i>	<i>-es</i>	<i>-es</i>	<i>džes</i>	<i>s</i>
Lithuanian	<i>-os</i>	<i>-a</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>dives</i>	<i>s</i>
Zakopane	<i>-os</i>	<i>-a</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>dives</i>	<i>s</i>
ESR	<i>-os</i>	<i>-a(s)</i>	<i>pes</i>	<i>-es</i>	<i>-es</i>	<i>d'ives</i>	<i>s</i>

Welsh	-os	-as	pes	-es	-es	dives	s
Finnish	-os	-as		-es	-es	diives	s
South Polish	-os	-a(s)	pes	-es	-es	ghives	s
Slovak	-os	-as		-es		dives	s
Radnitz*	-os	-as	pes	-es	-es	dives	s
Bugurdži	-os	-as	pes	-es	-es		s
West Bulgarian	-os	-as	pes	-es	-es		s
Sofia Erli	-os	-as	pes	-es		dijes	s
Sepečides	-os	-as	pes	-es			s
Rumeli	-os	-as	pes	-es	-es	dives	s

Noteworthy is first of all the actual pattern of *s*-reduction: Only occasionally do we still find an aspirated stage, which appears to be succeeded more or less throughout by zero –  $s > (h) > \emptyset$ . Figure 9 shows a group of dialects that are almost entirely consistent in reducing word-final *s* in all relevant items. A further class of dialects undergoes basically the same development, which however is resisted in the short forms of the present verb conjugation in the 2sg and 1pl, which have final *s*. The motivation to exempt these forms from the reduction of word-final *s* must be sought in the need to differentiate without ambiguity between present and subjunctive, or between future and present forms (the precise functional distribution being dialect-specific), and hence an avoidance of a mere opposition *-eha* : *-eh*, *-aha* : *-ah*, due to the weak status of the aspirated phoneme in marking meaningful syllable boundaries. The retention of *s* here is in other words a negative analogy to the long conjugation.<sup>3</sup> The only exception to this appears to be the Krajina Gurbet dialect, which possesses no long forms. A look at the geographical distribution of the dialects that undergo full reduction of word-final *s* – I shall return to the issue of geographical diffusion in more detail below – reveals a centre of diffusion in the southwestern Balkans (Kosovo, Macedonia, Albania, Montenegro) and stretching to the northwest (Krajina) to reach the southern Central dialects of Slovenia, eastern Austria, northern Hungary and western Slovakia.

Outside the areas of general *s*-reduction in word-final positions, we find several specific phenomena. The most common of those appears to be the loss of final *s* in the athematic ending (Greek) *-os*. There can be little doubt that this is an analogy to the nominal and adjectival masculine ending in *-o* in inherited (thematic) patterns, and so a levelling aiming at form: function consistency. This explains the lack of any clear centre of areal diffusion, although the development is not isolated either and we do find several different centres. Naturally, *-os > -o* occurs in those areas where final *s* is generally reduced. In addition we find it in Vlax, in the Sinti group (where it appears to be recent), and in the Polish and

3 For other types of morphophonological, dissimilatory processes in the southern Central dialects see Elšik/Hübschmannová/Šebková (1999).



North Russian varieties, while closely-related Lithuanian Romani retains forms in *s*. A further phenomenon is the reduction of *s* in the 3sg concord of the past conjugation; apart from the dialect with general reduction of final *s* this occurs in an area comprising the Baltic-Polish-Russian and northern Central dialects, as well as partly in Vlax. A less significant development finally is the loss of *s* in the reflexive pronoun, encountered outside the general reduction areas only in Vlax. We conclude from this, once more, that we are dealing with a complexity of phenomena involving the reduction of *s* in grammatical morphemes, with various analogies and simplification processes in operation alongside a general drift, confined to a particular geographical continuum, toward a straightforward phonological reduction.

Ferguson's (1990) "Spanish type" *s*-reduction foresees a change in both word-final and preconsonantal positions. Romani offers an interesting test-case: In the nominal paradigm, the accusative is etymologically identical to the general oblique which serves as a base for further case formation through attachment of consonant-initial suffixes. The m.sg accusative in *-es* therefore always appears in preconsonantal position when functioning as an oblique ending – dative *-eske*, ablative *-estar*, and so on. Figure 10 compares the forms for the (word-final) accusative and the etymologically parallel, preconsonantal oblique for a selection of dialects:

Figure 10: Word-final accusative and preconsonantal oblique

	acc m.sg	obl. m.sg
Montenegrinian Das	zero	h
Scutari Vlach	zero	h/s/zero
Hameln Sinti	zero	s
Istrian Sinti	zero	s
Roman	zero	s
Southern Central	zero	s
Romungri	zero	s
Krajina Gurbet	zero	s
Prishtina Gurbet	zero	s
Arlı	zero	s
Venetian Sinti	s/h	s
Kelderaš Finfiroiești	s/h	s/h
All other	s	s

Clearly, the phonological process in Romani cannot be said to have infiltrated preconsonantal positions beyond a very preliminary stage. The bulk of the dialects that show word-final reduction of *s* show no reduction at all in the preconsonantal position, while even those that do show a reduction here have one that is rather mild – retaining traces of the historical fricative in the form of an aspirate – rather than total. This can be added to the

evidence that suggests that we are dealing with a rather young phenomenon – other such evidence being the participation of dialects that are relative newcomers to the area that is the centre of diffusion in the southwestern Balkans, namely the Vlax dialects of the Gurbet, the Montenegrinian Das and the Scutari.

There is a further feature to be discussed in connection with the reduction of *s* in word-final position, namely the status of the remoteness marker or tense modifier in historical *-as*. Specifically, four possibilities must be considered. The first involves a form in *-as*, which can be found virtually throughout Romani varieties. This is likely then to be a conservative form, presumably the result of an enclitic, perhaps impersonal form of the copula in *\*as-* adjoined to the person-inflected tense forms to indicate temporal remoteness. A variation is encountered in some of the Sinti dialects, where the suffix is reduced to *-s* as a result of the appearance of initial stress and consequent overall syllable reduction. Both forms are represented in Figure 11 as retentions of *s*. The second type shows forms in *-ahi* in the southern Central dialects; this is indicated in Figure 11 by *h*. The third type shows full loss of the consonant and the presence only of a vocalic suffix in *-a*, represented in Figure 11 by "zero". Finally, some dialects, notably Arli, have lost the remoteness marker altogether and have developed analytic formations for the relevant functions. (The variety of Kelderash spoken by Taikon shows aspiration of the final consonant in the remoteness marker, indicated here as "h#").

Figure 11 shows firstly the absence of any inherent connection between *s > h > ∅* in the remoteness marker and the *s > h* development in morphological paradigms (Process A) – the two coincide only in the southwest Balkan varieties, but not in Sinti, Finnish or any of the Central dialects. On the other hand there is a consistent link between the reduction of *s* in word-final position (Process B) and the reduction of *s* in the remoteness marker, comprising both the southwest Balkan and the southern Central dialects. The outcome of this reduction differs however in each of the two areas. The remoteness marker drops its final consonant in the southwest Balkans, but assumes the form *-ahi* in the southern Central dialects. In principle, judging strictly by the shape of the formant, *-ahi* could have emerged as an attachment of a copula with an *h*-stem, which is common in this domain of dialects as a result of Process A; this is highly unlikely, however, since we must assume a Proto-Romani development that appears to be uniform in all other dialects. Moreover, the link between Process B (word-final reduction of *s*) and *-ahi* is sufficiently obvious. Elsewhere in the domain of Process B, aspirated consonants in word-final position seem to be a rather temporary state that is liable to further reduction, and we find "zero"-consonants in the remoteness marker. Only in Taikon's Kelderash do we find aspirated forms, alongside forms in *s* – hinting at the early stage of reduction. In the southern Central dialects the advanced stage will have faced two alternative pathways – the full reduction of the consonant, or a dissimilatory process – achieved by inserting a final vowel – allowing the survival of the aspirate.

Figure 11: The position of the remoteness (tense) marker

	instr	rem	<i>dives</i>
Sinti (General)	h	s	s
Finnish	h	s	s
South Polish	h	s	s
ESR	h	s	s
Zakopane	h	s	
Slovak	h	s	s
Radnitz	h	s	s
Cerhari	h	s	s
Roman	h	h	zero
Southern Central	h	h	zero
Romungri	h	h	zero
Montenegr. Das	h	zero	h
Arli	zero/j	–	zero
Prishtina Gurbet	s	zero	zero
Scutari Vlach	s/h	zero	zero
Taikon Kelderaš	s	s/h#	s
Krajina Gurbet	s	s	zero
Other (Welsh, Northeast, south- east Balkan, Vlax)	s	s	s

## 7. THE GEOGRAPHICAL SPREAD OF *S>H* PHENOMENA

We finally arrive at a consideration of the *s/h* alternation in the context of a geographical diffusion of structural innovations within Romani. Romani is one of the less-frequent cases of a European language that lacks a coherent spread over a particular territory. Moreover, Romani-speaking populations have a long history of migrations within Europe, and a significant proportion of them have traditionally engaged in commercial nomadism which in turn favours frequent changes of location. Finally, deliberate expulsions have often added their share to the factors that promote Romani migration. Franz Miklosich's early dialectological studies of Romani grant considerable attention to migration and its influence on the borrowed lexicon, attempting a classification of Romani dialects according to migration routes, while the works of Alexandre Paspati and Bernard Gilliat-Smith, among others, recognise occupational and settlement patterns as factors interacting with linguistic boundaries within Romani. Classifications of Romani dialects have since favoured large-scale groupings implying early ge-

netic connections, under which early splits from, or possibly within Early Romani and perhaps even Proto-Romani are understood. It is only very recently that studies in Romani dialect classification have begun to consider micro-phenomena, notably shared innovations and conservativisms in structures and classes of structures, as a basis for a differentiated description of individual isoglosses that separate or bound together varieties of the language. Inevitably their correlation with geographical spread is an issue under scrutiny.

We have so far identified several types of phenomena of *s/h* alternation, two of which may be regarded as "major-scale" – Processes A and B –, other as "minor-scale". The major issues are 1) The alternation in the copula and inflectional paradigms, and its analogical extension to interrogatives and determiners, itself a complex development involving a bundle of motivating factors (Process A), and 2) the phonological reduction of *s* in word-final position (Process B). Minor issues are the reduction of *s* in individual grammatical endings in loose connection both to each other and to the two major developments (e. g. the reduction of Greek-derived athematic m.sg inflection marker *-os* to *-o* by analogy to the inherited m.sg inflection marker, and so on). The "minor"-scale phenomena might be confined to particular dialects in isolation, or form areal isoglosses. Of the major issues, Process B – the phonological reduction in word-final position – has already been identified tentatively as a phenomenon with clear territorial spread. Process A remains the most challenging, as it assumes a starting point already inherited by Proto-Romani with variation at least in some of the forms (copula and verb inflection), with a subsequent spread to others (e. g. nominal inflection, interrogatives, determiners), part of which at least can be dated rather late, that is after the split following Early Romani, as it also affects a Slavic borrowing into Romani – *svako*>*sako*>*hako* 'every', or coincides with a compensatory renewal of a form through reinforcement by a German borrowing – *sa*>*ha*>*halauter* 'all'. Thus, Process A might be considered a combination of old inherited variants, preserved only by some varieties of the language, and later innovations which seem to have taken place more or less *in situ*.

Keeping these considerations in mind we turn to an evaluation of the geographical spread of *s/h* alternation phenomenon. Figure 12 is arranged to represent the territorial spread of Romani dialects from west to east and from north to south. Naturally, a linear presentation of a multidimensional spread has its limitations. In addition, decisions must be taken with regard to migratory dialects. The solutions adopted here include the placement of Finnish Romani in isolated position as a north-western dialect; adjoining the Italian Sinti dialects to the German and French Sinti-Manush, thus creating a north-south linearity; positioning Taikon's Kelderash speech recorded in Sweden in its area of origin in the northern Balkans; placing the Vlax dialects of the southern Balkans among the other southern Balkan dialects, to emphasise the fact that they share the relevant features with contiguous rather than historically related but non-contiguous Vlax dialects to the north:



A run through the issues dealt with in the preceding sections allows us the following observations:

*Process A*: There are two centres of diffusion. The first is situated in western-central Europe, its main projections being the Sinti-Manush and Central dialects. It also branches northwards via Scandinavia where it survives in the Finnish dialect, and southwards with the expansion of Sinti dialects from German-speaking territory to Italy. This geographical projection allows us to place the actual diffusion centre within Sinti, or German Romani, which links all participating branches. This geographical argument is supported by the structural features: Process A in this location assumes two types of forms, a "full" extent and a "partial" extent. The full extent features  $s > h$  in all functional paradigms including interrogatives and determiners as well as the copula in both tenses and in all persons. It is found in the centre itself, namely in those Sinti dialects that have remained in German-speaking territory (whilst outmigrant Sinti dialects show occasional forms in  $s$  indicating an uncompleted development). The partial extent shows  $s > h$  in the 3<sup>rd</sup> person copula present and in the inflection of nouns and verbs, partly also in the 3<sup>rd</sup> person copula past, only marginally in interrogatives and not at all in determiners or copula forms of the 1<sup>st</sup> and 2<sup>nd</sup> persons. It is found in the Central dialects and in Finnish Romani. The second diffusion centre is in the southwest Balkans, featuring a partial extent of  $s > h$  development. However, here we also find varieties of Arli with a double set for the copula throughout. The fact that Montenegrin Das, a Vlax dialect, shares this development speaks for an areal projection and inter-dialectal influence following the immigration of Vlax speakers. The two centres of diffusion for Process A are illustrated in Figure 13, which roughly represents their areal spread and location:

Figure: 13 Geographical diffusion of Process A ("full" and "partial")

	Finnish	Northeastern	
Welsh	<b>Sinti</b>	Central	Vlax
(Iberian)		southwest Balkan	southeast Balkan

Although a direct geographical continuum seems to meet the eye, linking the two centres, such an historical scenario need not be postulated. If we assume an historical carry-over from Proto-Romani of  $s/h$  variation in 1) the copula, and 2) the verb inflection (later triggering analogies within itself and in the nominal inflection, as proposed above), then it is plausible that some varieties of Early Romani will have retained this variation, while others eliminated it through simplification, specifically through the strict containment of  $h$ -forms

in modal expressions or the enclitic copula, abolishing them altogether in the inflectional paradigm. This development will have occurred before or parallel to the outwards migrations from the Balkans of the forerunners of Vlax, Iberian, Welsh, and the Northeastern dialects, as well as among the southeast Balkan varieties. Conversely, in at least two varieties the inherited variation remained, or led to the spread of *h*-forms at the cost of forms in *s* along the gradual hierarchical pattern described above. One such variety will have been the forerunner of Arli, its modern descendants themselves showing a variety of outcomes. The other variety will have been the forerunner of the dialect that settled in western-central Europe, and was eventually responsible for the western area of diffusion of Process A. The question that still demands an answer is, whether dialects that have undergone changes resulting from Process A diffusion must necessarily be seen as having immigrated into their current areas of settlements as a uniform group. This is essentially the question of whether the *s/h* isogloss justifies a genetic grouping within Romani. The internal evidence comfortably allows diffusion in situ within central Europe, with a generalisation of forms in *h* spreading – selectively, by category of affected items – from west to east. This scenario implies the emigration of the forerunner of Finnish Romani from western-central Europe prior to "Process A" reaching its full extent, and similarly it implies the outwards migration, from the centre of diffusion, of some of the Italian Sinti dialects at various stages during the evolution of "Process A". At the centre of diffusion itself, "Process A" reached its full extent in German Romani by the early 18<sup>th</sup> century at the latest, as documented by early attestations in sources such as the Waldheim Vocabulary and the Rotwelsch Grammar (see Matras 2000). External evidence reinforcing the impression of an areal diffusion of "Process A" are the features shared by the Northern branch of dialects, which include varieties that are not influenced by "Process A", such as Welsh, Iberian, and the Northeastern group (Polish, Baltic, Northrussian). The latter are known to have migrated into their present-day locations via German-speaking territories, and it is conceivable that they may have formed part of the original migrating population that left the Balkans and settled in central, western, and northern Europe from the late 14<sup>th</sup> century onwards. "Process A" may thus be assumed to have evolved after the split within this branch and settlement in the present locations.

There is however also external evidence (i. e. evidence that is not part of the *s/h* alternation itself) that points in another direction, namely that favouring a closer genetic affinity within the group of dialects that have undergone "Process A" to various degrees: There is a consistent correlation between the spread of "Process A" from its western diffusion centre, and the presence of *-l* in the 2sg past and copula conjugation, irrespective of the actual stem of the copula: thus Sinti *tu hal* 'you are', Slovak Romani *tu sal*. Elsewhere, the 2sg has merged with the 2pl in *-n*, rendering *tu san* as well as *tumen san* 'you(pl) are'. The form in

*-l* is thus inherited and conservative; its cognate can be found in the Domari 2sg ending *-r*. There are two possible explanations for this state of affairs. The first would be to assume that along with the diffusion of "Process A", other features were also diffused in a similar direction, among them the (re-)introduction of 2sg *-l*. In the dialect that served as the centre of diffusion, then, *-l* will have been an inherited conservatism, while in the other dialects affected by the diffusion *-l* will have been an innovation, albeit one that restored an original state of affairs. The alternative is to assume that the outwards migration from the Balkans from the 14<sup>th</sup> century onwards was all but uniform. It included the forerunners of Northern and Central dialects (as well as, in another direction, Vlax). Those will have shared some features, including innovations – e. g. the jotation of pronouns (*jov, joj, jon*) – but not others. The dominating group that settled in western-central Europe spoke varieties that included an active *s/h* alternation which in due course would lead to the emergence of "Process A", as well as a conservative feature – 2sg *-l*. The two explanations are thus not irreconcilable. More importantly, they can be reconciled with the clear areal pattern of diffusion encountered for "Process A".

*Process B*. Similarly, there are two centres of diffusion for the reduction of *s* in word-final position. The first is in the southwest Balkan dialects, represented in the sample by Arli and the Vlax dialects that have immigrated into the region. The second affects the southern Central dialects (with *-ahi* as a remoteness marker). There is also evidence that a contiguous group of southern Vlax dialects are partially affected by this development – the variation of *s/h* in word-final position in Taikon Kelderash; indeed, some word-final *s*-reduction in individual items such as the reflexive pronoun (*pes > pe*), the athematic m.sg ending or the 3sg past concord on the verb – extends even to "mainstream" Vlax, as represented by Kalderaš and even Lovari. The overall impression is one of three adjoining regions in which word-final reduction of *s* gains ground (see Figure 14), albeit partially to different extents and/or with different outcomes.

Figure 14: Geographical diffusion of Process B

	Finnish	Northeastern	
Welsh	Sinti	Central (north)	Vlax (north)
		Central (south)	Vlax (south)
(Iberian)		southwest Balkan	southeast Balkan

The fact that in-migrant Vlax dialects in the southwestern Balkans participate in "Process B", speaks both for a relatively late initiation of the sound change, and for inter-dialectal contacts that reach beyond the boundaries of tribes, religion, occupation and time of immi-



gration and settlement in an area. Although these boundaries surely do exist and have significant linguistic correlations – as can be seen in the sheer fact that in-migrating groups have resisted linguistic assimilation – the shared innovations observed here are clear testimony of the existence of an overall Romani speech community, despite internal diversity and lack of territorial coherence.

*Isolated reductions of final s.* In the areas that remain unaffected by "Process B", individual grammatical items show reduction of word-final *s*. The most common of those is the athematic m.sg ending, the reason for this having been explained above as an analogy to a regular nominal and adjectival m.sg ending in *-o*. No obvious explanation can be offered for the reduction of *s* in the reflexive pronoun or the 3sg past-tense concord marker, nor for the isolated cases of loss of *s* in the m.sg accusative marker in Sinti. As far as the reflexive pronoun is concerned, loss of *s* remains confined to regions adjoining the diffusion zones for "Process B" and so possibly we are dealing with interdialectal interference, in other words with areal diffusion, perhaps even with the seeds of a spread of "Process B" itself. As for the 3sg concord marker, we find a regional diffusion area in the Northeastern dialects, apparently unconnected to any of the other developments surveyed here, as well as a spread in areas adjoining the "Process B" zones, namely in Vlax. The inevitable conclusion is that, whilst wholesale, regular sound-changes of the type observed in "Process B"<sup>4</sup> may have some projection onto surrounding regions, one that is reflected in the reduction of final *s* in individual forms, such reduction may also be internally motivated, and in some case even lead to regional isoglosses.

## 8. CONCLUSION

I return to the three questions addressed in the introductory remarks:

- 1) As regard dialect classification, we have seen that the *s/h* alternation, though a conspicuous feature of variation among Romani dialects, involves several classes of phenomena, each containing in turn quite a few sub-phenomena. It is not possible nor advantageous to speak simply of *s*- and *h*-dialects, without naming specifically the nature of the development. This being said, however, we have been able to identify two major types of processes, labelled Process A and Process B for the purposes of this discussion. Participation in both these processes for individual dialects is not wholesale either, but gradual. Nonetheless, a statement on the extent to which a given dialect participates in

---

4 At the same time one must avoid the impression that even regular sound changes are entirely wholesale; see Elšik/Hübschmannová/Šebková (1999) for a discussion of how some words in the southern Central dialects remain unaffected by the loss of word-final *s*.

each of the processes, assuming this is the case, amounts to a classificatory statement on a class or even several classes of features displayed by this dialect. In this respect, *s/h* alternation can be used for dialect classification, provided of course the relevant phenomena are specified. The nature of such classificatory statement however is not in line with a notion of branching in tree diagrammes depicting linguistic-genetic affiliation. Rather, it is a statement on the extent to which dialects share innovations that might be a result of either divergence or diffusion. Indeed, throughout the above discussion a fairly solid case could be made for diffusion, while the arguments in favour of historically underlying divergence (a genetic affiliation of Sinti, Finnish, and Central dialects for example) remain largely speculative.

- 2) From this my conclusion is that areal diffusion and so the acquisition of innovations in situ plays a significant role in the diversity of Romani dialects, and that this role must be accentuated more strongly in dialectological studies of the language, which is indeed the current trend in Romani linguistics. This insight gained by the means of linguistic comparison suggests patterns of regular exchange and interaction among contiguous groups of Romani populations, which in turn implies historically more or less fixed areas of settlement. This calls for a possible revision of traditional notions of nomadism; for such ideas to be formulated, however, at first instance the opening of an interdisciplinary discussion is called for, comparing areal diffusion in the domains of culture and social organisation with linguistic findings.
- 3) There is no conclusive evidence that current dialect boundaries within Romani are either a replication of dialect boundaries within India or a reflection of boundaries among speaker populations of Proto-Romani. Pointing in this direction we might consider the split between western and eastern Apabhraṃśa in the verb ending of the 2sg (*h* in the former and *s* in the latter). The scenario suggested above for the roots of Process A assumes a spread from this form, as well as from the alternation in copula stems. In order to draw a connection with the Indian isogloss we would need to assume a part of the Proto-Romani population of eastern origin, maintaining this single eastern feature for a considerable period while assimilating in respect to virtually all other linguistic features. This scenario, while perhaps not absolutely impossible, cannot be argued for beyond pure speculation. We can nevertheless argue with some degree of confidence for an inherited variation, carried over into Proto-Romani from late MIA, serving as a trigger for some of the developments involving *s/h* alternation in present-day Romani, notably "Process A". In the phase that followed, each variety of Romani followed one of three possible courses – retention of variation, levelling in favour of *s*, and levelling in favour of *h*. Judging by the present distribution of Process A, it is very likely that a split

among dialects favouring one particular course of development emerged already in Early Romani. This does not, however, imply that the development was then contained by its initial boundaries; rather, subsequent areal diffusion, as argued above, will have led to the emergence of the state of affairs encountered today.

A final remark concerns a theoretical implication of the present discussion. The pattern of diffusion of *s>h* across grammatical forms and categories reveals a functional motivation for the change, namely the need for structural synchronisation of the grammatical apparatus. This synchronisation is of course only symbolic, in the sense that morphological forms that share a phonological feature – the phoneme *s* – through pure historical coincidence, are grouped together in a gradual, hierarchical progression of change that is not conditioned by articulatory factors. Only part of this development can be said to involve pure structural analogies – notably the shift to *h* in intervocalic position in the instrumental case and the long forms of present verb conjugation in the 2sg and 1pl. It is the link between these developments and *s>h* in the copula and interrogatives, which allows to argue the case for a wholesale grouping of grammatical categories on the basis of superficial structural similarities, and by implication, for a mental compartmentalisation of linguistic categories in which grammar and lexicon are kept apart.

## APPENDIX: LIST OF SOURCES FOR DIALECT DESCRIPTIONS CONSULTED

Arli (Kosovo)	Boretzky (1996)
Arli (Macedonia)	Kepeski/Jusuf (1980)
Auvergne Manush	Valet (1991)
Baukols (Italian) Sinti	Winstedt (1912)
Bohemian Sinti	von Sowa
Bugurdži	Boretzky (1993)
Cerhari	Mészáros (1976)
East Prussian Sinti	von Sowa (1893)
ESR (East Slovak Romani)	Hübschmannová/Šebková/Žigová (1991)
Finnish	Bourgeois (1911)
Hameln Sinti	Holzinger (1993)
Istrian Sinti	Soravia (1977)
Kalderaš	Boretzky (1994)
Krajina Gurbet	Jekh duj trin ...romanes
Lálere Sinte	Gilliat-Smith (1908)
Lithuanian	Māanušs (1997)
Lombard Sinti	Soravia (1977)
Lovari	Matras (1994)
Marburg Bible Sinti	Romanes-Arbeit Marburg e.V.
Molise	Ascoli (1865)
Montenegrenian Das	Matras, fieldwork
Moravian	von Sowa (1893)
North Russian	Wentzel (1980)
Piemontese Sinti	Soravia (1977) & Franzese (1985)
Polska Roma	Matras (1999)
Prishtina Gurbet	Boretzky (1986)
Radnitz	Puchmayer (1821)
Rhine Sinti	Gilliat-Smith (1907)
Roman	Halwachs (1998)
Romungri	Görög (1985)
Rüdiger's Sinti	Rüdiger (1782)
Rumeli	Paspati (1870)
Russian (North)	Sergievskii (1931)
Scutari Vlach	Mann (1933)
Sepečides	Cech/Heinschink (1997)
Sinti (Swabian?)	Finck (1903)
Slovak	Kalina (1882)
Sofia Erli	Boretzky (1998)
South Polish	Kopernicki (1930)
Southern Central	Elšik/Hübschmannová/Šebková (1999)
Taikon Kelderaš	Gjerdman/Ljungberg (1963)
Venetian Sinti	Soravia (1977)
Welsh	Sampson (1926)
Wesphalian Sinti	von Sowa (1893)
West Bulgarian	Minkov (1997)
Zakopane	Rozwadowski (1936)
Zigler Manush	Calvet/Delvoye/Labalette (1970)

## REFERENCES

- Ascoli, G.J. 1865 *Zigeunerisches*, Halle.
- Bakker, P. 1997 Athematic morphology in Romani: The borrowing of a borrowing pattern, in: Matras, Y./Bakker, P./Kyuchukov, H. (eds.) *The typology and dialectology of Romani*, Amsterdam: 1–21.
- Beames, J. 1872 (1970) *A comparative grammar of the Modern Aryan languages of India*, Delhi.
- Boretzky, N. 1986 Zur Sprache der Gurbet von Priština, *Giessener Hefte für Tsiganologie* 3, 1–4: 195–216.
- 1993 *Bugurdži. Deskriptiver und historische Abriß eines Romani-Dialekts*, Wiesbaden.
- 1994 *Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar*, Wiesbaden.
- 1995 Die Entwicklung der Kopula im Romani, *Grazer Linguistische Studien* 43, 1–50.
- 1996 Arli. Materialien zu einem südbalkanischen Romani-Dialekt, in: Halwachs D.W. (Hg.) *Romani I* (= Grazer Linguistische Studien 46): 1–30.
- 1998 Erli: Eine Bestandsaufnahme nach den Texten von Gilliat-Smith, *Studii Romani* V–VI: 122–160.
- Boretzky, N./Iglá, B. 1994 *Wörterbuch Romani-Deutsch-Englisch für den südosteuropäischen Raum*, Wiesbaden.
- Bourgeois, H. 1911 *Esquisse d'une grammaire du romani finlandais*, Torino.
- Bubeník, V. 1996 *The structure and development of Middle Indo-Aryan dialects*, Delhi.
- 1998 *A historical syntax of late Middle Indo-Aryan (Apabhraṃśa)*, Amsterdam.
- Calvet, G./Delvoje, F./Labalette, M. 1970 Abrégé grammatical de manuš, *Etudes Tsiganes* 16/1: 69–79.
- Cech, P./Heinschink, M.F. 1997 The dialect of the Basket-Weavers (Sepečides) of Izmir, in: Matras, Y./Bakker, P./Kyuchukov, H. (eds.) *The typology and dialectology of Romani*, Amsterdam: 159–181.
- Elšik, V./Hübschmannová, M./Šebková, H. 1999 The Southern Central (*ahi*-imperfect) Romani dialects of Slovakia and northern Hungary, in: Halwachs, D.W./Menz, F. (eds.) *Die Sprache der Roma. Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internationalen Kontext*, Klagenfurt: 277–390.
- Ferguson, Ch. 1990 From esses to aitches: identifying pathways of diachronic change, in: Croft, W./Denning, K./Denning, S. (eds.) *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75<sup>th</sup> birthday*, Amsterdam: 59–78.
- Finck, F.N. 1903 *Lehrbuch des Dialekts der deutschen Zigeuner*, Marburg.
- Franzese, S. 1985 *Il dialetto dei Sinti Piemontesi. Note grammaticali, Glossario*, Torino.
- Gilliat-Smith, B. 1907 The Gypsies of the Rhine province in 1902–1903, *JGLS new series* 1: 125–145.
- 1908 Lálere Sinte, *JGLS new series* 2: 12–14.

- Gjerdman, O./Ljungberg, E. 1963 *The Language of the Swedish Coppersmith Gipsy Johan Dimitri Taikon*, Uppsala.
- Görög, V. (ed.) 1985 *Tales of János Berki told in Gypsy and Hungarian*, Budapest.
- Halwachs, D.W. 1998 *Amaro vakeripe Roman hi. Unsere Sprache ist Roman. Texte, Glossar und Grammatik der burgenländischen Romani-Variante*, Klagenfurt.
- Holzinger, D. 1993 *Das Rómanes. Grammatik und Diskursanalyse der Sprache der Sinte*, Innsbruck.
- Hübschmannová, M./Šebková, H./Žigová, A. 1991 *Romsko-český a česko-romský kapesní slovník*, Praha.
- Kalina, A. 1882 *La langue des tziganes slovaques*, Posen.
- Kepeski, K./Jusuf, Š. 1980 *Romani gramatika*, Skopje.
- Kopernicki, I. 1930 *Textes tsiganes*, Krakow.
- Mann, St.E. 1933 Albanian Romani, *JGLS* third series 12: 1–32, 147–153.
- Masica, C.P. 1991 *The Indo-Aryan languages*, Cambridge.
- Matras, Y. 1994 *Untersuchungen zu Grammatik und Diskurs des Romanes. Dialekt der Kelderaša/Lovara*, Wiesbaden.
- 1999 The speech of the Polska Roma. Some highlighted features and their implications for Romani dialectology, *JGLS* fifth series 9: 1–27.
- 2000 Johan Rüdiger and the study of Romani in 18th century Germany, *JGLS* fifth series 9 (to appear)
- Mānuš, L. 1997 *Čigānu-Latviešu-Angļu etimoloģisla vārdnīca*, Riga.
- Mészáros, G. 1976 The Cerhāri Gypsy dialect, *Acta Orientalia Academiae Scientiarum Hung.* 30–3: 351–367.
- Miklosich, F. 1877 *Über die Mundarten und Wanderungen der Zigeuner Europa's VII*, Wien.
- Minkov, M. 1997 A concise grammar of West Bulgarian Romani, *JGLS* fifth series 7: 55–95.
- Paspati, A.G. 1870 (1973) *Etudes sur les Tschinghianés ou Bohémiens de l'empire Ottoman*, Osnabrück.
- Puchmayer, A.J. 1821 *Románi Čib, das ist Grammatik und Wörterbuch der Zigeuner Sprache*, Prag.
- Rozwadowski, J.M. 1936 *Wörterbuch des Zigeunerndialekts von Zakopane*, Kraków.
- Rüdiger, J.Chr.Chr. 1782 *Von der Sprache und Herkunft der Zigeuner aus Indien*, Hamburg (Reprint 1990).
- Sampson, J. 1926 *The dialect of the Gypsies of Wales*, Oxford.
- Sergievskaia, M.V. 1931 *Cyganskii jazyk*, Moskva.
- Soravia, G. 1977 *Dialetti degli Zingari Italiani*, Profilo dei dialetti italiani 22.
- Tagare, G.V. 1948 (reprint 1987) *Historical grammar of Apabhraṃśa*, Delhi.
- Turner, R.L. 1975 (1927) The phonetic weakness of terminational elements in Indo-Aryan, in: *Collected papers*, London: 291–300.
- Valet, J. 1991 Grammar of Manush as it is spoken in the Auvernhe, in: Bakker, P./Cortiade, M. (eds.) *In the margin of Romani. Gypsy languages in contact*, Amsterdam: 106–131.

- von Sowa, R.                    1893a Neue Materialien für den Dialekt der Zigeuner Deutschlands, *Zeitschrift der deutschen morgenländischen Gesellschaft* 47: 450–465.
- 1893b Die mährische Mundart der Romsprache, *Jahresbericht der ersten Deutschen K.K. Gymn. in Brünn* 1892–93: 1–19.
- 1888 Die Mundart der ostpreussischen Zigeuner, *Zeitschrift für Völkerpsychologie* 18: 82–93.
- 1889 Die Mundart der westfälischen Zigeuner, *Zeitschrift für Völkerpsychologie* 19: 192–204.
- Wentzel, T.                    1980 Die Zigeunersprache (Nordrussischer Dialekt), Leipzig.
- Winstedt, E.O.                1912 O Bovedantuna: A tale in French Romani. Communicated by Augustus E. John, *JGLS new series* 5: 204–218.
- Woolner, A.C.                1924 Studies in Romani philology III: The verb substantive, *JGLS third series* 3: 180–184.