

Some gradient dimensions of marginal vowels: The example of Romani in contact with Hungarian

MÁRTON A. BALÓ* 

ELTE Research Centre for Linguistics, Institute for General and Hungarian Linguistics, Hungary

Received: March 20, 2025 • Revised manuscript received: April 12, 2026 • Accepted: April 21, 2026

© 2026 The Author(s)



ABSTRACT

This paper aims to make a contribution to the theoretical evaluation of marginal vowels, that is, vowels whose phonemic status may be called into question due to, among other things, their low functional load and hence weak contrastivity, their borrowed origin or their unpredictable appearance. The example taken is Romani, a New Indo-Aryan language which has been widely spoken in Europe and, throughout its history, has come into contact with various languages, including Hungarian. The paper introduces new dimensions and attempts to expand existing ones along which the gradient nature of the sounds of a language can be interpreted, providing clues about the varying degrees to which different sounds may play a role within the sound system.

KEYWORDS

marginal vowels, Romani, Hungarian, contact, phonemicity

1. INTRODUCTION

Romani, a New Indo-Aryan language spoken outside the Indian subcontinent, is said to have a fairly stable inventory of core vowels; on the other hand, it has also been in more or less intensive contact with many different languages since its speakers left the Indian Sprachbund, which has inevitably left its mark, among other things, on the vowel system. Naturally, the vowel systems of the other New Indo-Aryan languages also underwent change, which resulted in the six (e.g. in Nepali) to thirteen (e.g. in Sinhalese) or possibly even more (e.g. Kashmiri) vowels of

* Corresponding author. E-mail: balo@nytud.hu

the present day. However, the development of these vowel systems seems to have been much less influenced by other languages (for more detail, cf. [Masica 1991](#), 113–115). Up to the appearance of Romani in Europe, its phonological system seems to have remained largely unaffected by other languages, and even Byzantine Greek, which had a lasting influence on the syntax of Romani, left its phonology almost intact, perhaps its only consequential influence being the loss of vowel length. The languages Romani came into contact with after its dispersal in Europe, many of which are very different from Romani typologically, enriched its inventory of sounds; however, despite the easily perceptible phonetic differences between Romani dialects spoken wide apart, the phonemicity of the new or modified sounds is often ambiguous when they come under closer scrutiny.

A marginal phoneme in this context is broadly understood as any phoneme which is not part of the basic Romani stock and, therefore, has given rise to various different attempts to categorise and interpret it from being an outlier and being swept under the carpet through taking part in ambiguous allophonic relationships to carrying different degrees of functional load in contrastive relationships. In this paper, the focus is on vowels which can be considered marginal in this sense.

In what follows, I try to underscore the gradient nature of the phonemicity of sounds through the example of Romani in contact with Hungarian. First, I give an overview of the evolution of the five-vowel system of Romani from Old Indo-Aryan (Section 2), and then describe the current state of affairs as regards the newly emerged vowels, focusing on the Romani varieties in contact with Hungarian, chiefly South Central Romani and Vlax, while also touching upon some other Romani dialects (Section 3). After discussing some of the potential pitfalls and explanations (Section 4), I propound five dimensions along which the gradience of sounds may be interpreted in general, and which may help handle marginal phonemes better than categorical models (Section 5).

2. THE EVOLUTION OF THE VOWEL SYSTEM OF ROMANI

Present-day Romani stands out from the rest of the New Indo-Aryan languages in terms of its vowel system, being the only one which contains five vowels, as opposed to a minimum of six vowels in the subcontinent (for an overview of the vowel systems of New Indo-Aryan languages, see [Masica 1991](#), 109–113). The five-vowel system appears to originate from Late Proto-Romani¹ ([Baló 2020](#), 120; [Matras 2002](#), 58), which, in turn, goes back to the five short vowels of the Middle Indo-Aryan vowel system ([Beníšek 2020](#), 28). The Middle Indo-Aryan vowel system was a symmetrical one of five short and five long vowels ([Masica 1991](#), 184–185). In this respect, Romani is apparently more conservative than the other New Indo-Aryan languages, which developed additional, phonemic vowels.

The Old Indo-Aryan vowel system had three short vowels, three corresponding long vowels and seven vocalic segments: a, i, u, ā, ī, ū, ē, ō, āi, āu, ɳ, ɳ̄, ɳ̅. The vocalic segments are comprised of the diphthongs āi and āu; the phonetically purely monophthongal ē and ō, which were,

¹[Beníšek \(2020, 18\)](#) introduces the term Late Proto-Romani, designating what is called Early Romani by [Matras \(2002\)](#), to be more “in line with practices common in historical linguistics,” and defines it as the “common forerunner of all Romani dialects ... which was followed by its split into individual dialects”.

however, considered diphthongs because of certain sandhi rules in Sanskrit; and the syllabic consonants ṛ, ṛ̥ and ḷ (for a more detailed overview, see Masica 1991, 163–166).

As Late Proto-Romani is the form that is suggested to have been spoken in the Byzantine Empire at the beginning of the second millennium (Beníšek 2020, 13), which coincides with the approximate date of the beginning of what is referred to as the New Indo-Aryan stage in the history of Indo-Aryan (Masica 1991, 50–51), it will be both convenient and satisfactory to review the changes which the Old Indo-Aryan vowels listed above underwent to reach the five-vowel system of Late Proto-Romani. A thorough overview of the phonological changes shared with other Indo-Aryan languages, as well as the conservative features of and the innovations shown by Romani are found in Matras (2002, 30–41). A summary of the changes and illustrative examples can be found in Table 1.

As Matras (2002, 34) explains, the short vowel /a/ was retained where it historically preceded a consonant cluster and changed to /e/ before a single consonant, while the /a/ representing the original masculine singular inflectional ending changed to /o/. Thus, there are at least three different, widespread outcomes of /a/, and, in addition, there are also sporadic instances of Old Indo-Aryan medial /a/ appearing as /i/ or /o/ in Romani (see Table 1; cf. Matras 2002, 34). The short vowel /i/ was either retained or changed to /e/ before /r/ (Michael Beníšek, personal communication, 16 Feb 2024; see also Oslon & Kozhanov 2025), while /u/ was generally retained across the board (see Table 1; cf. also Matras 2002, 39), although there may be some fluctuation (e.g. *upre/opre* ‘up, above’, from Old Indo-Aryan *uppari*).

Table 1. Vowel changes from Old Indo-Aryan to Romani

	Old Indo-Aryan	Romani
a > a	<i>gharma</i> ‘heat’	<i>kham</i> ‘sun’
a > e	<i>ghara</i> ‘house’	<i>kher</i> ‘house’
a > o	<i>kṣīṇa</i> ‘worn away’	<i>khino</i> ‘tired’
u > u	<i>mukha</i> ‘mouth’	<i>muj</i> ‘mouth’
i > i	<i>cikka</i> ‘gummy matter in eyes’	<i>čik</i> ‘mud’
i > e	<i>śiras</i> ‘head’	<i>šero</i> ‘head’
ā > a	<i>bhrātṛ</i> ‘brother’	<i>phral</i> ‘brother’
ī > i	<i>jīvati</i> ‘is alive’	<i>dživel</i> ‘lives’
ū > u	<i>dūra</i> ‘distant’	<i>dur</i> ‘far’
ē > e	<i>devatā</i> ‘divinity’	<i>devel</i> ‘god’
ō > o	<i>lohita</i> ‘red’	<i>lolo</i> ‘red’
āi > e	n/a	
āu > o	<i>gāura</i> ‘white’	<i>goro</i> ‘peasant’
ṛ > u	<i>pr̥cchāti</i> ‘asks’	<i>phučel/pučel</i> ‘asks’
ṛ > i	<i>śṛṅga</i> ‘horn’	<i>šing</i> ‘horn’

The long vowels /ā, ī, ū/ were not affected by these changes and kept their quality, which is in line with the assumption that length was only lost at a later stage in the history of Romani, under the influence of Armenian or Greek, as Boretzky & Igla (2004, 34) suggest. This loss of length in Romani also affected /ē/ and /ō/. Besides this, another, earlier source of the new short vowels /e/ and /o/, which may have contributed to their becoming potential target vowels of other vowels shifting to /e/ and /o/ (see above), was a general shortening process which affected all Old Indo-Aryan long vowels before consonant clusters, making them positional allophones (Masica 1991, 182). Another overall change affected the Old Indo-Aryan diphthongs /āi/ and /āu/: they became /ē/ and /ō/, respectively (Masica 1991, 168), which were later shortened to /e/ and /o/ in Romani.

A general change which concerns all final vowels in most New Indo-Aryan languages is their apocope; this is usually mentioned in connection with short vowels, but as long final vowels had been shortened, they were also lost (Masica 1991, 183, 188).

As for the Old Indo-Aryan vocalic liquids ṛ, ṝ, ṛ̥, they were lost in Middle Indo-Aryan and were replaced by /a, i, u/ (Masica 1991, 167). In Romani in general, the syllabic ṛ seems to have become /u/ after labials and /i/ elsewhere (Osion & Kozhanov 2025). The syllabic ṛ̥ only appeared in certain forms of one single, not very common verbal root, *klp* (Whitney 1879, 10), meaning ‘order, arrange, manage’, and has no trace in Romani. Likewise, there is no continuation of the long ṝ in Romani: it only appeared in accusative and genitive (and, in the case of the neuter declension, also the nominative and vocative) plural forms of the paradigm of a comparatively limited nominal class, “almost entirely composed of derivative nouns” formed with the agentive suffix *tr̥*, “and a few nouns of relationship not made with that suffix” (Whitney 1879, 123). These nouns of relationship included *mātr̥* ‘mother’, *pit̥r̥* ‘father’ and *svāsr̥* ‘sister’; however, the Romani words with these meanings, *dad*, *daj* and *phen* go back to other Old Indo-Aryan roots.

3. THE CURRENT STATE OF AFFAIRS

Although these five short vowels have been said to be fairly stable, they have undergone certain modification processes in the various dialects of Romani and may be realised somewhat differently, often under the influence of the contact languages, which have also furnished Romani with new vowels. The status of both the modified and the new vowels can be rather fluid, as we will see below.

3.1. Borrowed vowels

The front vowels /y/ and /ø/ (and their long counterparts) belong to the group of new vowels and appear in Romani dialects in contact with Hungarian, Turkish or Albanian, among others: Turkish *köprü* ‘bridge’ > Xoraxani Romani *köprüs*, Albanian *qytet* ‘town’ > Mečkari Romani *tjüteti* (Baló 2020, 122), Swedish *hög* ‘high’ > Finnish Romani *hüög*, Swedish *byk* ‘laundry’ > Finnish Romani *büükæ* (cf. Valtonen 1968, 1972, cited by Granqvist 1999, 48, 2002, 64).²

²The Finnish data come from Granqvist (1999, 48, 2002, 64), who cites Valtonen (1968, 1972). Valtonen (1968) also suggests that /y/ and /ø/ were added through contact with Hungarian.

As Elšík, Hübschmannová & Šebková (1999, 309) explain, they are delabialised and adapted to the conservative vowel system in what he calls Northern South Central Romani varieties, e.g. R. *bégéši* < H. *bögös* ‘double bassist’ (Vlčany, Slovakia), R. *fitin-* ‘heat V’ < H. *füt* (Zohor, Slovakia). This also applies to Vend Romani, where “Hungarian front labialised vowels are generally adapted by delabialisation” (Bodnárová 2015, 93). The process is also attested in Vlax Romani dialects in Hungary, e.g. R. *kisebo* ‘threshold’ < H. *küszöb* (Vekerdi 1985, 138–139, collected in Esztergom, Hungary). However, this is far from being always the case. In another tale in the same volume, told by a South Central Romani speaker, the same word appears in the form *küsobo*, without unrounding (Vekerdi 1985, 290–291, collected in Varsány, Hungary). Examples of the retention of roundedness are not infrequent, irrespective of the specific dialect: H. *völegény* ‘bridegroom’ > R. *völegénjo* (Vekerdi 1985, 58–59, Vlax Romani, collected in Kispeszt, Hungary), H. *gyógyfű* ‘herb’ > R. *djödjjűv-* (Vekerdi 1985, 276–277, South Central Romani, collected in Varsány, Hungary), H. *gyümölcs* ‘fruit’ > R. *djümölčo* (Vekerdi 1985, 312–313, South Central Romani, collected in Varsány, Hungary). The unpredictability of the delabialisation of /y/ and /ø/ is evident from the fact that there may even be intra-speaker variation, and even within the same sentence: *trin eltenje/trin öltönje gāda* ‘three sets of clothing’ < H. *öltöny* (Vekerdi 1985, 144–145, Vlax Romani, collected in Esztergom, Hungary). The unrounded form only appears once, and it is the first occurrence of the word; this sentence is followed by more instances of the retention of the rounded vowel by the same speaker.

The handling of borrowed vowels only occurring in loanwords is further complicated by the fact that Romani has applied different, so-called xenoclitic (formerly referred to as athematic) morphological patterns to loanwords, which “signal their foreign origin” (see Elšík 2020, 168–170 for more details). The “morphological encoding of the origin of lexemes”, called “chronological compartmentalisation” by Elšík & Matras (2006, 324), marks the words out as foreign. Labialisation, if present, is just an additional, phonological feature, as these vowels never appear in the inherited lexicon, which has its own, dedicated, so-called oikoclitic (formerly referred to as thematic) morphology.

3.2. On the borderline between borrowed vowels and modified vowels

As Elšík, Hübschmannová & Šebková (1999, 308) report, the South Central Romani varieties spoken in Vlčany and Selice (Slovakia) took over the short open front vowel /æ/ along with loanwords from the local Hungarian dialects, e.g. *kæzdin-* < H. *kezd* ‘begin’. This vowel corresponds to Standard Hungarian /e/ and could have been adapted to Romani /e/, as it has happened in various other varieties. However, the opposite thing happened: the borrowed vowel /æ/ also began to be used in various inherited words,³ e.g. *khær* ‘house’ instead of *kher*. According to Elšík, Hübschmannová & Šebková (1999, 308), “the phonological status of the vowel /æ/ remains unclear: it seems that at least in a majority of the inherited words it is in allophonic relationship with /e/,” perhaps because it often occurs in a stressed syllable before a liquid, but there are examples in which it is followed by a non-liquid consonant, such as *dæš*

³The inherited lexicon contains Indo-Aryan, Iranian and Armenian words, as well as a significant number of Byzantine Greek loanwords shared by all Romani dialects. The borrowed lexicon refers to loanwords acquired by the various dialects independently, after their dispersal across Europe beginning in the fourteenth century (cf. Scala 2020 for a thorough discussion of the composition of the lexicon of Romani; and see also Matras 2002, 14–25 and Beníšek 2020).

‘ten’ instead of *deš*. Another source of /æ/ is also listed in the Vlčany variety, namely the unstressed low middle /a/ in *zijaen* ‘damage’ instead of *zijan*.

Another somewhat ambiguous allophonic situation is described by Bodnárová (2015, 89–92) in her description of the vowel inventory of another South Central Romani variety, the Vend dialect spoken in Kisbajom (Hungary). Both Hungarian [ɛ] and the short mid front unrounded vowel [e], a prominent feature of the Hungarian dialect spoken in the region (cf. Juhász 2001, 267, 271, 275; Kiss 2006, 520–522), are taken over. When they are used in words borrowed from Hungarian, the Hungarian distribution is copied. However, their distribution in the inherited lexicon is more complex. Firstly, there is a tendency for complementary distribution, with the mid variant used in the final syllable and the mid-open variant used elsewhere, e.g. *edej* ‘here’ ~ [edej]. Secondly, open [æ], which was also borrowed from the local Hungarian dialect (cf. Király 2005, 35, cited by Bodnárová 2015, 91), and open-mid [ɛ] are in free variation. Neither the allophonic, nor the free variation lends phonemes to the phoneme inventory of the dialect, which contains only one short /ɛ/. In addition, the reason why original Romani /e/ ends up as one or another vowel in certain cases is not clear, cf. e.g. *efta* ‘seven’ ~ [ɛfta] but *enja* ‘nine’ ~ [ɛna]⁴ (the examples are from Bodnárová 2015, 99); here, both initial vowels go back to the same Greek, and, by extension, to the same Romani vowel.

Gabor, a Northern Vlach Romani variety closely related to Kalderash is reported to have borrowed /ɛ/ from Hungarian, while also retaining the mid front vowel /e/ (Gardner & Gardner 2008, 166–167). The vowel /ɛ/ has also infiltrated the inherited component of the lexicon, replacing various, morphologically definable strings containing the vowel /a/. In the nominal system, some examples are the nominative plural suffix *-ja*, e.g. Kalderash⁵ *šurja* ‘knives’ ~ Gabor *šure*; the sequence *-ja* in the nominative plural suffix *-urja*, e.g. Kalderash *vasurja* ‘dishes’ ~ Gabor *vasure*; the whole sequence *-aja* in feminine singular demonstratives, e.g. Kalderash *kadaja/kodoja* ‘this/that’ ~ Gabor *kade/kode*; and even the nominative plural suffix *-a*, e.g. Kalderash *šošoja* ‘rabbits’ ~ Gabor *šošoje*. Gardner & Gardner (2008) also mention minimal pairs which are the result of this process, such as *gaže* ‘non-Romani women’ (~ Kalderash *gažja*) versus *gaže* ‘non-Romani men’ or *medve* ‘bear’ versus *medve* ‘bears’ and *kečke* ‘goat’ versus *kečke* ‘goats’.⁶ The Gabor verbal system is also affected by the existence of both /e/ and /ɛ/. The intransitive marker *-(j)av-* changes to *-ev-*, e.g. Kalderash *šiljav-* ‘be cold’ ~ Gabor *šilev-*, and the second person singular imperfect suffix is contracted, e.g. *avex* ‘come 2SG IMP’ instead of **avexax* (cf. Kalderash *avesas*), contrasting with the second person singular present form *avex* ‘come 2SG PR’ (cf. Kalderash *aves*). The more open back vowel /ɔ/, also borrowed from Hungarian, replaces the *-ava* and *-ova* endings of masculine singular demonstratives, e.g. Kalderash *kadava/kodova* ‘this/that’ ~ Gabor *kado/kodɔ*, and the *-ava-* sequence in the first person singular imperfect ending: *avɔx* ‘come 1SG IMP’ instead of **avavax* (cf. Kalderash *avavas*); however, in this case, no minimal pairs are documented.

⁴Short /ɛ/ and long /e:/ are treated as separate phonemes, similarly to Hungarian, the contact language; see more about vowel length in Section 3.4.

⁵The Kalderash data are taken from Boretzky (1994).

⁶The words *medve* and *kečke* must have been borrowed from Hungarian; they are documented in the forms *medva* and *kečka* in other Romani varieties influenced by Hungarian, such as Lovari (cf. Cech & Heinschink 2002).

3.3. Modified vowels

A process that has affected the quality of the front vowels in various Romani dialects is centralisation (Baló 2020, 125–126). This usually happens to varieties in contact with languages that have either /i/ or /ə/ in their inventory of phonemes or allophones. An example of the process involving a Kalderash dialect and Romanian is described in detail by Boretzky (1991). Besides their presence in Romanian loanwords, /i/ is replaced by /i/ after the sibilants /s, z, ʃ, ʒ, t͡s/ and the uvular /ʁ/ in the inherited lexicon, while /e/ is replaced by /ə/ after the same sounds, but also before the same sounds and after the velars /g, k, kʰ/. In his discussion of the phonemic status of the centralised vowels, Boretzky (1991, 10–11) concludes that “the whole problem hinges on the uncertainties of phonetic realisation rather than on the existence of shared or distinct environments” but finds no minimal pairs. The environments in which the two pairs of vowels appear are only similar, and only if the inherited and borrowed parts of the lexicon are treated as one unit, cf. *gîndil* ‘thinks’ (< Romanian *gîndi*) ~ *ginel* ‘counts’ (< Old Indo-Aryan *gr̥náyati*) and *mərətîl* ‘marries’ (< Romanian *mărita*) and *merel* ‘dies’ (< Old Indo-Aryan *maratē*). Based on this, it seems more reasonable to consider them as allophones of the same phoneme in complementary distribution. Other Romani varieties provide us with no better clue, either: in North Russian Romani, centralisation appears to be stress-dependent; in various other dialects, it may spread to other positions, including varieties in which it takes place across the board, or, alternatively, results in free variation (cf. Baló 2020, 125–126).

3.4. Long vowels

Generally, no phonemic status has been attributed to vowel length in Romani (Baló 2020, 126), in particular because historically long vowels merged with their short counterparts in Proto-Romani (Beníšek 2020, 28), most probably under the influence of Greek, or perhaps Armenian (see Boretzky & Iglá 2004, 33–34). Some of the reasons why long vowels reappeared in various Romani dialects listed by Baló (2020, 126–128) are compensatory lengthening and the influence of contact languages such as Lithuanian, Latvian, Finnish, Hungarian, Slovak and Slovene. Indeed, vowel length in South Central Romani must have developed in contact with Hungarian and not earlier (cf. Elšík, Hübschmannová & Šebková 1999, 310–311; Bodnárová & Wiedner 2015, 158–159; Elšík 2022, 206; Bodnárová & Wiedner 2022, 221–222).

The presence of long vowels in Hungarian loanwords was instrumental in the spread of vowel length to the inherited lexicon, the potential processes of which are spelt out by Elšík (2022, 204–212), who focuses his description on Selice Romani, a South Central Romani variety. In line with Bodnárová & Wiedner (2022, 222), who partly relate the development of vowel length in South Central Romani to stress, he also emphasises the role of pretonic lengthening (e.g. *khangerí* ‘church’ > **khangēri* > *khángēri*). He also adds several dimensions, such as the openness of the syllable, the presence of an onset, the sonority of the coda, the labiality of the following onset, the height of the vowel and the length of the word, along which pretonic lengthening took place in a gradient way; hence, not every pretonic vowel is long. On the other hand, long vowels also occur in positions which were not originally pretonic. This is explained by analogical extension which also results in the reduction of allomorphic variation, cf. plural *akhorá* ‘nuts’ > *ákhora* (and not **ákhōra*), in analogy to singular *akhór* ‘nut’ > **ákhór* > *ákhor*. The analogical extension of vowel length in possessive pronouns and the multi-functional suffix

-ahi in Vend Romani is described by Bodnárová & Wiedner (2015). The latter process resulted in a functional split, with *-ahi* as the imperfective marker and *-āhi* as the irrealis marker. However, the marker is attached to different stems in its different functions; therefore, this differentiation will not yield minimal pairs (see examples (1) and (2), taken from Bodnárová & Wiedner 2015, 166):

- (1) ker-eh-ahi
STEM.PRS-2SG-IMP
'you were doing; you would do'
- (2) kerd'-al-āhi
STEM.PF-2SG-IRR
'you would have done'

One more source of long vowels is the coalescence or contraction of adjacent heterosyllabic short vowels (Elšík 2022, 210, and see also Bodnárová & Wiedner 2015, 159). Compensatory vowel lengthening accompanied by the loss of /j/ or /v/ is also mentioned by Bodnárová & Wiedner (2015) and Bodnárová & Wiedner (2022), as, for example, in *duj* > *dū* 'two' or *avrí* 'out' > *āri*. Baló (2020, 128) adds that vowel lengthening may not necessarily be accompanied by the deletion of /v/, cf. Finnish Romani *džūli/cūli/jūli* 'woman' but also Romungro *žūvli* (< *džuvli*), while Elšík (2022, 211–212) suggests that in certain instances, such as the perfective stem, the elision of /v/ actually preceded lengthening, e.g. *tav-* 'cook' ~ **tav-d-* PF > **ta-d-* > *tā-d-*; *khuv-* 'plait, weave' ~ **khuv-d-* PF > **khu-d-* > *khū-d-*.

3.5. Processes involving long vowels

Short vowels and their corresponding long counterparts may take part in allomorphic alternation (Elšík 2022, 193–204); some of these vowel shortening and lengthening processes are also accompanied by some other, complementary process.

In the domain of verbs, the third-person singular and second/third-person plural future and imperfect forms of verbs are obligatorily “syncopated”, that is, the suffixes *-el* and *-en* take the forms *-l* and *-n*, respectively, which is accompanied by the shortening of the root vowel of certain (but not all) verbs (see Table 2; the data are from Elšík 2022, 194).

The long vowel of a few feminine nouns is shortened in their inflectional paradigm, which co-occurs with the appearance of a more complete form: NOM SG *sū* 'needle' ~ NOM PL *suvja* (containing the original word-final consonant), NOM SG *čūci* 'breast' ~ NOM PL *čučča*

Table 2. Second and third person singular and plural forms of the verb *tāv-* 'cook' in Selice Romani

	2SG	3SG	2/3PL
present	<i>tāves</i>	<i>tāvel</i>	<i>tāven</i>
future	<i>tāvesa</i>	<i>tavla</i>	<i>tavna</i>
imperfect	<i>tāvesahi</i>	<i>tavlahi</i>	<i>tavnahi</i>

(containing a geminate consonant). Derivational suffixes also frequently trigger vowel shortening, particularly in nominal and adverbial bases.⁷

The distinction between long vowel and short vowel forms in adjectives was already noted by Elšík, Hübschmannová & Šebková (1999, 311), and it was described in more detail and more accurately by Elšík (2022, 196–198). It involves the shortening of the lexical long vowel of adjectives in attributive position (see examples (3) and (4), taken from Elšík 2022, 197, and Table 3, based on Bodnárová & Wiedner 2022, 228):

- (3) O lačo manuš āja.
the good-M-SG-NOM man-SG-NOM come-PF-3SG
The good man came.
- (4) O manuš lačo hi.
the man-SG-NOM good-M-SG-NOM COP-PRS-3SG
The man is good.

The phenomenon was further investigated by Bodnárová & Wiedner (2022), who also included the long and reduced forms of the possessive pronouns (e.g. *mīro/mro* ‘my’). They found that the distinction is motivated by the different discourse-pragmatic functions of the attributive and predicative positions, with the long vowel forms fulfilling rhematic and the short vowel (or reduced) forms fulfilling thematic roles. They also add that this is part of a broader tendency to employ phonologically more complex forms in independent position and less complex forms in dependent position, citing a few adjectives which contain a geminate consonant when used predicatively, e.g. *gulo* ~ *gullo* ‘sweet’. It is worth noting that this phenomenon also seems to be a gradient one both geographically and in terms of the lexicon. It is very widespread in two- and three-syllable adjectives containing an open penultimate syllable but less so in two-syllable adjectives containing a closed first syllable. Among the varieties Bodnárová & Wiedner (2022) look at, there is less alternation in the ones spoken in the north of Hungary, and more in the south of Hungary.⁸

Table 3. Adjectives in attributive and predicative positions in South Central Romani

	Attributive	Meaning	Predicative	Meaning
Masculine singular	<i>bar-o kher</i>	big house	<i>o kher bār-o hi</i>	the house is big
Feminine singular	<i>bar-i bār</i>	big garden	<i>i bār bār-i hi</i>	the garden is big
Plural	<i>bar-e gāva</i>	big villages	<i>o gāva bār-e hi</i>	the villages are big

⁷It has to be noted that “although long vowels most commonly occur in open syllables, they are also well attested in closed syllables, before both word-final and word-internal codas” (Elšík 2022, 190). An example of a verb is *nāš-* ‘run’: 3SG PRES *nāšel*, 3SG FUT *nāšla*, 3SG IMP *nāšlahi*; an example of a noun is NOM SG *hēv* ‘hole’ ~ NOM PL *hēvja*. Short vowels in open syllables are abundant.

⁸The reason why the Romani variety spoken in Nógrádszakál contains so many adjectives in which the vowel, contrary to all expectations, is not long in any position remains to be accounted for.

Lengthening processes include the addition of the perfective suffix, which affect several verbs (see also above), and some nominal and adverbial derivations, affecting a few primary nominal bases with a monosyllabic root. Adpositions and the corresponding adverbs contain short and long vowels, respectively, and there even exist two minimal pairs: *anglal* “from the front of” versus *ānglal* “in/to the front” and *palal* “from the back of” versus *pālal* “in/to the back” (Elšík 2022, 189–190). There are also a few more minor instances of vowel lengthening, such as the ones affecting some deictic forms or the transitive verbal suffix *-ar-*.

3.6. Interim summary

In Section 2, the Romani system of five short vowels was shown to have evolved, through various sound changes, from the original Old Indo-Aryan vowel system, which consisted of thirteen elements. Even among these thirteen elements we can find vowels that may be considered marginal: the use of the vocalic segments $\bar{r}$ and $\bar{l}$ was rather restricted and their contrastive nature is questionable.

The problem of marginality became even more evident in the ensuing enrichment of the Romani vowel system, which was described in detail in Section 3, and all the examples demonstrate that both borrowed and modified vowels pose questions regarding their status, and their categorisation is problematic.

New vowels such as /y/ and /ø/ appeared in various Romani dialects in contact with different languages via the borrowing of new lexical items, and, although loanwords are originally treated morphologically differently from the inherited lexicon, the case is not so clear-cut (see Section 4). The open-mid and near open front vowels /ɛ/ and /æ/, of Hungarian origin, appeared across the board in Romani varieties as distant as the Vend and Gabor dialects, alongside the inherited close-mid front vowel /e/, in the latter interfering even in the morphological system. In Romani dialects in contact with languages already containing them, central /i/ or /ə/ have also started to spread as part of the centralisation process of original Romani /i/ and /e/. Finally, long vowels reappeared in varieties influenced by contact languages in which vowel length had already been present, and they also began to play a role in several morphophonological processes, such as the syncopation of the imperfect forms of certain verbs or the attributive versus predicative positioning of adjectives.

4. DISCUSSION

The vowels /y/ and /ø/ were introduced as part of a foreign stratum in the lexicon (cf. Hall 2013, 237–239). In Romani dialects in contact with Hungarian, minimal pairs also arose (e.g. *šero* ‘head’ versus *šöro* ‘beer’ < H. *sör* in sample HU-009 in Matras & Elšík 2001–2016), and, according to Hall (2013, 228), the existence of minimal pairs is usually considered a robust argument for the contrastivity of two sounds. However, as shown in Section 3.1, retention of the labial feature is far from being the only strategy, and the vowel of the very same word may be delabialised to *šero* in the speech of other speakers of the same dialect recorded elsewhere (Elšík 2009, 297). In addition, there has been interaction between the oikoclitic and the xenoclitic compartment, mentioned in Section 3.1, and the strict morphological split has shown signs of weakening (see Elšík & Matras 2006, 327–332; Baló 2021). As Hall (2013, 246–250) explains,

applying a stratified or polysystemic model to the phonology of a language can be appropriate if there are no lexical items which would cross over, but, in the Romani case, it is exactly the borrowed elements, potentially containing front rounded vowels, to which the oikoclititic morphology may be extended. While relegating these vowels to a separate stratum would be convenient, as they only appear in Hungarian loanwords, there are also several counterarguments: firstly, their appearance cannot be established with certainty; secondly, even if they are labialised, there are many other borrowed words which do not contain them; and, thirdly, the morphological behaviour of the lexical items containing them may deviate from the xenoclititic patterns.

If we take a closer look at the vowel /*ɛ*/ in the inherited lexicon of the Gabor dialect, we see that it evolved from a variety of segments which are, in most cases, more complex than just a single vowel. This specific case is perhaps best interpreted through the notion of “functional load” as described by Hall (2013, 250–252), that is, in terms of the number of minimal pairs and information loss if the contrast were to disappear, although the lack of enough data does not allow us to calculate either. Following Wedel, Kaplan & Jackson (2013), who find that phonemes with a high functional load are less likely to merge over time, we may surmise that /*ɛ*/ and the contrasts it now takes part in are here to stay. Interestingly, Gardner & Gardner (2008, 156) do not include /*ɛ*/ in the vowel chart, only /*e*/, but they later refer to them as distinct phonemes (Gardner & Gardner 2008, 166). One of the things that makes this case unique and the difference unlikely to be neutralised is the fact that, as opposed to the Vend case, it is unlikely that this would be a case of the complete adoption of a distinction already existing in the Hungarian contact variety (cf. e.g. Juhász 2001, 279, 296; Kiss 2006, 522, 524); rather, /*ɛ*/ was borrowed and /*e*/ was retained.

Another, different way of interpreting the functionality of a specific sound is the extent to which it is functionalised grammatically. Matras (2002, 60), for instance, reports that “morphological endings such as the terminations of demonstratives (Southern Central and Welsh Romani) or plural markers may show consistent lengthening,” which is not unlike multiple exponence (Harris 2017), where a single feature or function is expressed through more than one realisation within a word. The differences in vowel length accompanying the position of adjectives (see Section 3.5) represent a similar phenomenon: the position of the adjective before the noun is a sufficient marker of the attributive function; nevertheless, it is enhanced by the shortening of the vowel in the adjective. The lengthening of the vowel /*a*/ in the suffix *-ahi* (see Section 3.4) is also akin to this: its position after the present or the perfective stem is already a signal of its function, which is further reinforced by the difference in length. We may say that the distribution of these forms of the functionalisation of vowel length is predictable, and they are never the sole realisations of the function or feature they mark; they are bound up with other, morphosyntactic properties of the containing element.

5. GRADIENCE

The notion of gradience (see Hall 2013, 259–262) has already been touched upon a couple of times. Although we do not have a sufficient amount of data in the form of a robust corpus to be able to use a model such as the Probabilistic Phonological Relationship Model (PPRM, Hall 2012), we can still attempt to make some more general inferences. The interpretation of the

marginal vowels discussed above can be made along various dimensions, all of which are of a gradient nature.

1. Naturally, the first dimension is the “cline” formed by five kinds of contrast, ranging from allophonic to contrastive segments, as formulated by Goldsmith (1995, 10–11; referred to as the “Cline of Contrast” by Hall 2013, 260).⁹ The only vowel that can be considered truly contrastive is /ɛ/ in the Gabor dialect (see Section 3.2). In other varieties, however, it is considered an allophone, along with /æ/. The central vowels /i/ and /ə/ of the Kalderash dialect (see Section 3.3) only appear in environments similar to the ones in which the front vowels /y/ and /e/ appear; therefore, they tend towards the allophonic end of the scale. Based on minimal pairs such as the one mentioned above (*šēro* ‘head’ ~ *šōro* ‘beer’), the front vowels /y/ and /ø/ may be considered phonemic. In their case, however, the second dimension also has to be taken into consideration (Figure 1).
2. The presence of the vowel in the inherited lexicon constitutes the second dimension. The front vowels /y/ and /ø/, if they are not delabialised, appear in borrowed words only. In Romani varieties in contact with Hungarian, even their presence is not predictable, and it seems to be of a gradient nature both geographically and temporally, although this is not something that can easily be proved based on the data currently at our disposal. In any case, these two vowels, if they occur at all, they only occur in the borrowed lexicon, and never spill over into the inherited lexicon.¹⁰ Others, such as /ɛ/, /æ/ and the centralised vowels do, but most often only allophonically (Figure 2).
3. The third dimension may be defined as faithfulness to the original form, in terms of both the inherited and the borrowed lexicon. While the delabialisation of the front vowels /y/ and /ø/ increases faithfulness to the Romani sound system, it decreases faithfulness to the Hungarian sound system. Conversely, the borrowing and spread of the front vowels /æ/ and /ɛ/, as well as the centralisation of /i/ and /e/ decrease faithfulness to the inherited sound system, which

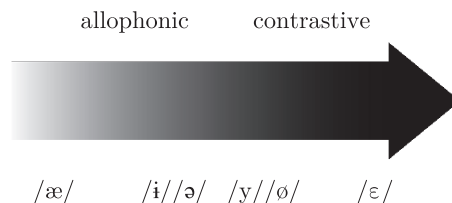


Figure 1. The dimension of allophony-contrastivity

⁹Marginal contrasts may also be better accommodated by models which take into account redundancy and frequency effects (Pierrehumbert 2012) as well as the fact that mental representations of phonological forms are extremely detailed (Pierrehumbert 2016, 48).

¹⁰According to Granqvist (1999, 2002), based on Valtonen (1968), /y/ and /ø/, albeit rarely, may appear in the inherited vocabulary in Finnish Romani, e.g. *phüüli* ‘widow’ instead of *phibli/pivli/pili* in other dialects, or *thüöoli* ‘cigarette’ instead of *thuvali* in other dialects.

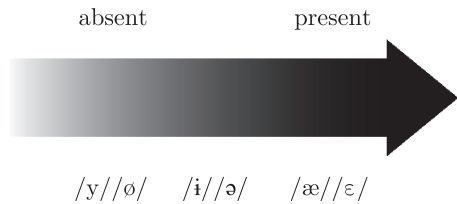


Figure 2. The dimension of absence or presence in the inherited lexicon

is further reduced if they begin to appear in more and more environments or they become more and more phonemic (Figure 3).

4. The degree of grammatical functionalisation makes up the fourth dimension. In our sample, with the exception of /ɛ/, vowel length appears to be the only feature which results in varying degrees of grammatical functionalisation; as Elšík (2022, 193) puts it, “vowel quantity takes part in a number of synchronic morphophonological processes.” The data presented in Section 3.5 show that there are several morphological operations which trigger either vowel shortening or vowel lengthening, resulting in phonological allomorphs. On the other hand, the complete functionalisation of vowel length would be akin to base modification as described, for example, by Haspelmath & Sims (2010, 35–39), such as palatalisation of the final consonant in the plural formation of certain Albanian nominal classes (*murg* ‘monk’ > *murgj* ‘monks’); gemination of a medial consonant in causative verb derivation in Standard Arabic (*darasa* ‘learn’ > *darrasa* ‘teach’); or vowel shortening in intransitive verb formation in Hindi (*ma:r-* ‘kill’ > *mar-* ‘die’). Such examples are few and far between in Romani: one potential candidate is the geographically restricted pattern of the first and second person singular possessive pronouns in Nógrádszakál Romani, where the reduced variants (*mro*, *tro*) are used non-emphatically, while the long variants (*mīro*, *fīro*) are used emphatically: *mro kher* ‘my house’ versus *mīro kher* ‘my house (not someone else’s)’ (Bodnárová & Wiedner 2022, 238) (Figure 4).
5. Finally, a fifth dimension may also be added, as the type and intensity of the expansion of the vowel system vary significantly across dialects, with some dialects incorporating more vowels or to a larger extent, while others may be more conservative in terms of the number of additional or modified vowels, the changes they trigger and the morphophonological processes they take part in.

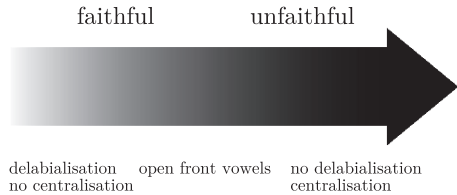


Figure 3. The dimension of faithfulness to the inherited vowel system

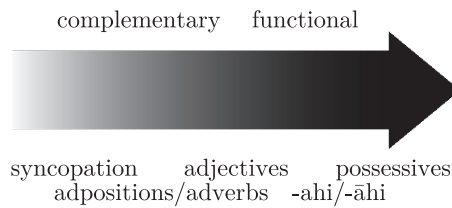


Figure 4. The dimension of the grammatical functionalisation of vowel length (selected examples)

6. FINAL REMARKS

As [Hall \(2012, 2\)](#) puts it, “one of the primary criteria for determining phonological relationships is that of predictability of distribution,” adding that there may be cases in which two sounds are “mostly predictable, with some degree of unpredictability, or mostly unpredictable, with some degree of predictability,” potentially wavering on the boundary of allophone and phoneme. Through the example of various Romani dialects in contact with Hungarian, and also citing some other Romani varieties in contact with other languages, I have attempted to complement the existing theoretical assumptions about the notion of gradience in phonology, such as the cline of different kinds of contrast ([Goldsmith 1995](#)) or the Gradient Phonemicity Hypothesis ([Ferragne et al. 2011](#)) by adding new dimensions or expanding the existing ones along which vowels can be interpreted to better explain the fact that the vowels, or indeed, the sounds used in a particular language can play smaller or larger roles within the sound system, and while allophony and contrastivity can be significant, they are not the only factors to be taken into consideration.

REFERENCES

- Baló, Márton A. 2020. Romani phonology. In Y. Matras and A. Tenser (eds.) *The Palgrave handbook of Romani language and linguistics*. Cham: Palgrave Macmillan. 119–153.
- Baló, Márton A. 2021. Variation in the nominal morphology of Northern Vlax Romani. *Word Structure* 14(1). 25–58.
- Beníšek, Michael. 2020. The historical origins of Romani. In Y. Matras and A. Tenser (eds.) *The Palgrave handbook of Romani language and linguistics*. Cham: Palgrave Macmillan. 13–47.
- Bodnárová, Zuzana. 2015. *Vend Romani: grammatical description and sociolinguistic situation of the so-called Vend dialects of Romani*. Doctoral dissertation. Prague: Charles University.
- Bodnárová, Zuzana and Jakob Wiedner. 2015. A comparative study of verbal particles in varieties of Vend Romani. *Romani Studies* 5 25(2). 197–216.
- Bodnárová, Zuzana and Jakob Wiedner. 2022. Development of phonemic length alternation in adjectives and possessive pronouns in South Central Romani. *Romani Studies* 5 32(2). 221–245.
- Boretzky, Norbert. 1991. Contact induced sound change. *Diachronica* 8. 1–16.
- Boretzky, Norbert. 1994. *Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar*. Wiesbaden: Harrassowitz.

- Boretzky, Norbert and Birgit Igla. 2004. *Kommentierter Dialektatlas des Romani*. Wiesbaden: Harrassowitz.
- Cech, Petra and Mozes F. Heinschink. 2002. In *Wörterbuch der Romani Variante der Österreichischen Lovara. Lovari-Deutsch-Englisch. Arbeitsbericht 3A des Romani-Projekts*, hrsgg. von Dieter W. Halwachs. Vienna: Verein Romano Centro.
- Goldsmith, John. 1995. Phonological theory. In J. A. Goldsmith (ed.) *The handbook of phonological theory*. Cambridge, MA: Blackwell. 1–23.
- Gardner, David J. and Sarolta A. Gardner. 2008. A provisional phonology of Gabor Romani. *Romani Studies* 5 18(2). 155–199.
- Elšik, Viktor. 2009. Loanwords in Selice Romani, an Indo-Aryan language of Slovakia. In M. Haspelmath and U. Tadmor (eds.) *Loanwords in the World's languages. A comparative handbook*. Berlin: De Gruyter Mouton. 260–303.
- Elšik, Viktor. 2020. Romani morphology. In Y. Matras and A. Tenser (eds.) *The Palgrave handbook of Romani language and linguistics*. Cham: Palgrave Macmillan. 155–186.
- Elšik, Viktor. 2022. Vowel length in Selice Romani: Phonology, morphophonology, and diachrony. *Romani Studies* 5 32(2). 185–219.
- Elšik, Viktor, Milena Hübschmannová and Hana Šebková. 1999. The Southern central (ahi-imperfect) Romani dialects of Slovakia and Northern Hungary. In D. W. Halwachs and F. Menz (eds.) *Die Sprache der Roma. Perspektiven der Romani-Forschung im interdisziplinären und internationalen Kontext*. Klagenfurt: Drava. 277–390.
- Elšik, Viktor and Yaron Matras. 2006. *Markedness and language change: the Romani sample*. The Hague: Mouton de Gruyter.
- Ferragne, Emmanuel, Nathalie Bedoin, Véronique Boulenger and François Pellegrino. 2011. The perception of a derived contrast in Scottish English. In W. S. Lee and E. Zee (eds.) *Proceedings of the 17th International Congress of phonetic sciences*. Hong Kong: City University of Hong Kong. 667–670.
- Granqvist, Kimmo. 1999. Notes on Finnish Romani phonology. *Grazer Linguistische Studien* 51. 47–63.
- Granqvist, Kimmo. 2002. Finnish Romani phonology and dialect geography. *SKY Journal of Linguistics* 15. 61–83.
- Hall, Kathleen Currie. 2012. Phonological relationships: A probabilistic model. *McGill Working Papers in Linguistics* 22.
- Hall, Kathleen Currie. 2013. A typology of intermediate phonological relationships. *The Linguistic Review* 30(2). 215–275.
- Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.
- Haspelmath, Martin and Andrea D. Sims. 2010. *Understanding morphology*, 2nd edn. London: Hodder Education.
- Juhász, Dezső. 2001. A magyar nyelvjárások területi egységei [The territorial units of Hungarian dialects]. In J. Kiss (ed.) *Magyar dialektológia [Hungarian dialectology]*. Budapest: Osiris. 262–324.
- Kiss, Jenő. 2006. Nyelvjárások, regionális nyelvváltozatok [Dialects, regional vernaculars]. In F. Kiefer (ed.) *Magyar nyelv [Hungarian language]*. Budapest: Akadémiai Kiadó. 517–548.
- Király, Lajos. 2005. *Somogy-zalai nyelvátlasz [Linguistic atlas of Somogy and Zala]*. Budapest: Magyar Nyelvtudományi Társaság Kiadványai.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron and Viktor Elšik. 2001–2016. *Romani Morpho-syntax (RMS) database*. Manchester: University of Manchester. Available at: <https://romani.dch.phil-fak.uni-koeln.de/>.

- Osłon, Mikhail Vladimirovich and Kirill Kozhanov. 2025. Этимологический словарь цыганского языка [Etymological dictionary of the Romani language]. Manuscript. Available at: <http://rromanes.org/pub/%D0%AD%D0%A1%D0%A6%D0%AF.pdf>.
- Pierrehumbert, Janet B. 2012. The dynamic lexicon. In A. C. Cohn, C. Fougeron and M. K. Huffman (eds.) *The Oxford handbook of laboratory phonology*. Oxford: Oxford University Press. 173–183.
- Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus Episodic. *Annual Review of Linguistics* 2: 33–52.
- Scala, Andrea. 2020. Romani lexicon. In Y. Matras and A. Tenser (eds.) *The Palgrave handbook of Romani language and linguistics*. Cham: Palgrave Macmillan. 85–117.
- Valtonen, Pertti. 1968. Suomen mustalaiskielen kehitys eri aikoina tehtyjen muistiinpanojen valossa [The development of the Finnish Romani language in the light of records made at different times]. Licentiate Thesis. Helsinki: University of Helsinki.
- Valtonen, Pertti. 1972. Suomen mustalaiskielen etymologinen sanakirja [Etymological Dictionary of the Finnish Romani language] (Tietolipas 69). Helsinki: Suomalaisen Kirjallisuuden Seura.
- Vekerdi, József. 1985. Cigány nyelvjárás népmesék [Gypsy vernacular tales]. Debrecen: Kossuth Lajos Tudományegyetem.
- Wedel, Andrew, Abby Kaplan and Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179–186.
- Whitney, William Dwight. 1879. *A Sanskrit grammar*. Leipzig: Breitkopf and Hartel. London: Trübner & Co.