

## Analogical extension of vowel length in Vend Romani

**Zuzana Bodnárová**

Dpt. of Linguistics, Charles University Prague &  
Dpt. of Finnish, Finno-Ugrian and Scandinavian  
Studies, University of Helsinki  
zuza.bodnarova@gmail.com

**Jakob Wiedner**

Department of Linguistics and Scandinavian  
Studies, University of Oslo  
j.a.p.wiedner@iln.uio.no

**Abstract:** Distinctive vowel length has been only recently re-introduced into the Romani varieties that have been in intimate contact with the various languages of Europe exhibiting vowel quantity. This article describes the process of analogical extension that accounts for certain intra-dialectal variation of vowel length found within the South Central Romani dialect group. The emergence of vowel length by means of this process is demonstrated by the example of the possessive pronouns and the remoteness suffix of Vend Romani, a variety spoken in Western Hungary. This analysis also discusses the phonological and semantic constraints of the examined instances of analogical change.

**Keywords:** analogy; extension; vowel length; markedness; Vend Romani

### 1. Introduction

At the beginning of the second millennium, owing to the prolonged contact that Romani had with either Armenian or Greek, or with other languages of the Balkans, distinctive vowel length has been lost in Romani (Miklosich 1872–1880, IX: 24; Sowa 1887, 18; Boretzky & Igla 2004, 34). However, length opposition was re-introduced into various Romani dialects that were influenced by post-Greek contact languages that distinguish vowel length, such as Hungarian, Slovak, Czech, German, Welsh and Finnish (see Matras 2002, 59–60). These Romani dialects have thus developed distinct patterns of length distribution. By contrast, the patterns of vowel lengthening in these dialects display several similarities that appear to have resulted from the similar processes of vowel lengthening found in the respective contact languages (Boretzky & Igla 1993, 36). In that regard, the presence

of distinctive vowel length in the present-day Romani dialects is considered to be an areal contact feature (Matras 2002, 59–60).

In the past, scholars have focussed primarily on the inter-dialectal variation of vowel length, while variation that occurs within the individual dialect groups of Romani has not received much attention thus far. Elšík et al. (1999, 309–313) also note that the intra-dialectal variation of length distinction requires further research, basing this on their observation that the distribution of long vowels in the varieties of the South Central dialect group is not equal. The present paper accounts for such variation within the South Central dialect group by describing an innovative process of lengthening that occurs in Vend Romani, a South Central Romani variety spoken in Western Hungary (see Vekerdi 1984; Bodnárová 2013), while being absent in other varieties of the same dialect group. It will be argued that the introduction of long vowels in Vend Romani resulted, in certain cases, from the process of analogical extension, triggered either by compounding or contraction. Analogical extension is understood here to be a type of linguistic change involving a less common form that has extended and therefore become the more common form (see Hock 1986, 238–279; Anttila 2003; Blevins & Blevins 2009). We will demonstrate this particular change by examining certain possessive pronouns and a part of the verb inflection paradigm. We will also attempt to identify the constraints on vowel lengthening by considering the phonological and semantic contexts in which this pattern extension arises.

The source of our data on Vend Romani includes transcriptions of elicited and spontaneous language data that have been recorded by the first author of the article in 2011 as a part of the project *Linguistic Atlas of Central Romani* 2011–13 (P406/11/0818, The Czech Science Foundation) at the Charles University in Prague. The sample comprises ca. 350 000 word-form tokens. The Vend Romani texts collected by Rézműves (2006) have been consulted only as secondary data. The data on other South Central Romani varieties are drawn from the same project.

## 2. Vowel length and processes of vowel lengthening

In situations where a language is under permanent influence of a more prestigious language, as is the case with linguistic minorities, it can be observed that the phonological system of the minority language is adjusted to the phonology of the majority language (Matras 2009, 222–225). Within such contact-induced changes the adaption of vowel length seems to occur more likely than adjustment of vowel quality (*ibid.*, 231–232). Therefore,

we may assume that the introduction of vowel length into Vend Romani has been triggered by prolonged contact with Hungarian, a language that has vowel length opposition. As a result, Vend Romani has ten independent vowel units, five short (1a) and five long (1b), in addition to the front-rounded short (2a) and long (2b) vowel pairs that are borrowed from Hungarian:

- (1) a. *a* /a/, *e* /ɛ/, *i* /i/, *o* /o/, *u* /u/  
 b. *á* /aː/, *é* /ɛː/, *í* /iː/, *ó* /oː/, *ú* /uː/
- (2) a. *ö* /ø/, *ü* /y/  
 b. *ő* /øː/, *ű* /yː/

Vowel length is a distinctive phonological feature of Vend Romani that can be established on the basis of minimal pairs such as the following: *ov* ‘be.IMP.2SG’ vs. *óv* ‘he’; *sapano* ‘wet’ vs. *sápano* ‘snake’s, of snake’; *meg* (verbal particle denoting perfective aspect) vs. *még* ‘still, yet’; *khul* ‘crap’ vs. *khúl* ‘s/he weaves’; *phral* ‘brother’ vs. *phrál* ‘s/he opens’; *so* ‘what’ vs. *só* ‘what kind of’; *asál* ‘s/he makes so. laugh’ vs. *ásal* ‘s/he laughs’, etc. Long vowels may occur in any position in a word (e.g., *ičutno* ‘yesterday’s’, *adí* ‘today’, *paramísi* ‘story, tale’), and a word may contain more than one long vowel (e.g., *pékíbe* ‘cake’, *khírínel* ‘s/he shouts’, *lákjáhi* ‘s/he would have found’). The distribution of long vowels is not bound to the position of stress. The fact that long vowels and stress are independent of one another is exemplified in table 1 where stress is marked with bold.

**Table 1:** Long vowel in stressed and unstressed syllables

Number of syllables	Long vowel	
	Stressed syllable	Unstressed syllable
1	<b>bár</b> ‘garden’	—
2	<b>mánuš</b> ‘human’	már <b>á</b> ‘I will beat’
3	dand <b>é</b> rel ‘s/he bites’	lákj <b>á</b> hi ‘s/he would have found’

In Vend Romani, as well as in Romani in general, long vowels have been introduced through contraction and compensatory lengthening (see Matras 2002, 60), e.g., *žá* < \**žava* ‘I will go’, cf. *žav* ‘I go’; *dí* < \**dives* ‘day’. Vowels have also become lengthened in open syllables in pre-tonic position (*idem.*), e.g., CVCV \**ka.lo* > \**ká.lo* ‘black’, cf. \**šud.ro* ‘black’. This

development represents an earlier stage of the language when the stress fell on the final syllable. More recently, the stress has shifted to the penultimate syllable in Vend Romani, resulting in the coincidence of long vowels with stressed open syllables, e.g., CVCV *tá.ha* ‘tomorrow’ (cf. CVCCV *tík.no* ‘small’), CVCVC *má.nuš* ‘human’ (cf. CVCCVC *čum.ňik* ‘whip’), CVCVCV *mo.mé.li* ‘candle’ (cf. CVCVCCV *či.rik.li* ‘bird’). Vowel length is thus to a large extent predictable in case of the inherited lexicon. Hungarian loan words, on the other hand, are always adapted together with their vowel length, e.g., *té.vé* < Hung. *tévé* ‘television’, *pu.ló.ve.ri* < Hung. *pulóver* ‘pullover’, *pa.lo.ta* < Hung. *palota* ‘palace’. In these loans, the initial syllable is stressed.

Elšík et al. (1999, 311) report that vowel length became morphologically relevant in South Central Romani, where the inherited adjectives and the polysyllabic possessive pronouns become shortened when used attributively, e.g., *amaro verda* ‘our car’, cf. predicative adjective use *o verda amáro hi* ‘the car is ours’. They (*ibid.*, 312) also noticed that the vowel length of the base form is generally preserved in derivations as well as throughout the inflectional paradigm of the word, e.g., *pé.kav* ‘I bake’ vs. *pék.łum* ‘I baked’, but not in *phé.nav* ‘I say’ vs. *phen.dum* ‘I said’, etc. However, a thorough analysis is needed to determine in which environment the vowels become lexicalized.

In the following two sections, we will describe a process in which vowels become long through analogical change. It will be shown that the source of such lengthening is found in either open-syllable lengthening or contraction-based/compensatory lengthening.

### 3. Analogical extension of vowel length in possessive pronouns

Table 2 illustrates the set of possessive pronouns found in Vend Romani, where we can distinguish between monosyllabic (*mr-*, *tr-*, *pr-*) and polysyllabic pronouns (*les-kr-* ~ *les-ker-*, *la-kr-* ~ *la-ker-*, *amar-*, *tumar-*, *pumar-*, *len-gr-* ~ *len-ger-*).<sup>1</sup>

<sup>1</sup> Vend Romani employs the reflexive possessive pronouns *pr-* ‘his/her (own)’ and *pumar-* ‘they (own)’ in sentences with two coreferential noun phrases. For instance, the sentence *óv kámel pra lumňa* ‘he loves his (own) wife’ involves the reflexive possessive *pr-*, while the sentence *me kámv leskra lumňa* ‘I love his wife’ the possessive pronoun *leskr-*.

**Table 2:** Possessive pronouns of Vend Romani

1SG	mr-	1PL	amar-
2SG	tr-	2PL	tumar-
3SG.REFL	pr-	3PL.REFL	pumar-
3SG.M	les-kr- ~ les-ker-	3PL	len-gr- ~ len-ger-
3SG.F	la-kr- ~ la-ker-		

Possessive pronouns take the adjectival endings that are represented by the short vowels M.SG *-o*, F.SG *-i* and PL *-e* in the nominative (3), and M.SG *-e*, F.SG *-a* and PL *-e* in the oblique<sup>2</sup> (4).

- (3) mr-o murš ‘my-M.SG husband’  
 mr-i čhaj ‘my-F.SG daughter’  
 mr-e khéra ‘my-PL houses’
- (4) mr-e muršes- ‘my-OBL.M.SG husband.OBL’  
 mr-a čha- ‘my-OBL.F.SG daughter.OBL’  
 mr-e kheren- ‘my-OBL.PL houses.OBL’

In Vend Romani most of the singular gender markers of monosyllabic possessive pronouns have developed a long vowel counterpart, which is only employed when the head noun is *daj* ‘mother’ or *dad* ‘father’, e.g., *mr-í daj* (< \**mr-i daj*) ‘my-F.SG mother’, *mr-ó dad* (< \**mr-o dad*) ‘my-M.SG father’. The same innovation occurs in the South Central varieties spoken in Páty (Pest county) and Versend (Baranya county).<sup>3</sup> By contrast, the

<sup>2</sup> The oblique forms of possessive pronouns are employed when the head noun is in cases other than nominative. The oblique form of the noun corresponds to the accusative case (except for the masculine singular form where the final /s/ is preserved, e.g., *dad-e* ‘father-ACC’, cf. *dad-es-* ‘father-OBL.M.SG’), and it is used to mark the animate direct object. Further cases are formed from the oblique stem of the noun through the attachment of the respective case markers, e.g., *d-a-tar* ‘mother-OBL.F.SG-ABL; from mother’, *dad-es-ke* ‘father-OBL.M.SG-DAT; to father’.

<sup>3</sup> A similar innovation has been attested in some North Central varieties in southeastern Slovakia (Turňa and Abov regions), as well as in the Transcarpathian Ukraine (Michael Beníšek, personal communication, 12 July 2014). For instance, in Moldava nad Bodvou Romani (Turňa), the commonly used possessive pronouns are *mír-* ~ *mr-* ‘my’ and *tír-* ~ *tr-* ‘your’, but only the contracted forms *mr-* ‘my’ and *tr-* ‘your’ occur next to the nouns *daj* ‘mother’ and *dad* ‘father’. The gender/number marker is short in the former (e.g. *mír-i čhaj* ~ *mr-i čhaj* ‘my-F.SG daughter’, *mír-o sastro* ~ *mr-o*

gender/number markers are always short in other South Central varieties, including the “vendic” (also called Vend Romani) varieties spoken in Austria and Slovenia, e.g., *mi daj* in Mátraverebély Romani (Hungary), *mi dáj* in Dunajská Streda Romani (Slovakia), *mri daj* in Oberwart Romani (Austria), or *mri dej* in Dolina Romani (Slovenia).

The lengthening of vowels in the monosyllabic possessive pronouns *mr-V* ‘my’, *tr-V* ‘your’ *pr-V* ‘his/her own’ can most likely be traced to the commonly used phrases of ‘my/your/his/her own mother’ and ‘my/your/his/her own father’, which have become compounds in Vend Romani, e.g., *\*mri + daj < \*mri daj*, *\*mro + dad < \*mro dad*. Similar compounding is found in the Spanish-Romani mixed language called “Caló”, in which the nouns *minday* and *chinday*, both denoting ‘mother’, have most likely emerged from the compounds *\*minři daj* ‘my mother’ and *\*tinři daj* ‘your mother’, respectively (Boretzky 1998, 106). This development is particularly interesting because compounding is not a productive means of word formation neither in Vend Romani nor in Romani in general (Matras 2002, 119).

In the newly emerged compounds, the gender markers of pronouns have become located in open pre-tonic syllables and have therefore undergone lengthening, e.g., *\*mrí.daj < \*mri daj*, *\*mró.dad < \*mro dad*. Following the shift in stress patterns, long vowels are now found in stressed syllables, e.g., *mrí.daj < \*mrí.daj*, *mró.dad < \*mró.dad*. While it still inflects for gender, the pronoun constituent of the compound is strongly bound to the head noun, e.g., *mr-í + daj* ‘my-F.SG + mother’ (long vowel), but cf. *mr-i moštóvni daj* ‘my-F.SG step-mother’ (short vowel).

Subsequently, the occurrence of long vowels has become analogically extended to the feminine oblique forms of the monosyllabic pronouns when followed by the nouns *daj* ‘mother’ and *dad* ‘father’ (table 3).<sup>4</sup>

In the oblique, the stress falls on the oblique suffix of the noun instead of the pronoun, i.e., *d-a- < \*daj-a-* ‘mother-OBL.F.SG’, *dad-es-* ‘father-OBL.M.SG’. It has been mentioned in the previous section that vowel length is generally maintained in the inflectional forms of words. According to this, one would expect that the feminine oblique form of the compound

---

*sastro* ‘my-M.SG father-in-law’), and long in the latter case (e.g. *mr-í daj* ‘my-F.SG mother’, *mr-á da-* ‘my-OBL.F.SG mother.OBL.F.SG’, *mr-ó dad* ‘my-M.SG father’). This change has not affected the third person singular reflexive pronoun, which has only the non-contracted (i.e., polysyllabic) form in the local dialect, i.e., *peskr-*.

<sup>4</sup> This does not hold for the irregular vocative form *mri dáj* ‘my.F.SG.NOM mother.VOC’, which consists of the nominative form of the pronoun and the apocopated form of the noun (*dáj < \*dáj-e* ‘mother.VOC’, cf. *daj*, ‘mother’).

**Table 3:** Innovative vowel lengthening in the monosyllabic possessive pronouns

	‘my ~ your ~ his/her own mother’	‘my ~ your ~ his/her own father’
NOM	mr- <b>í</b> +daj ~ tr- <b>í</b> +daj ~ pr- <b>í</b> +daj	mr- <b>ó</b> +dad ~ tr- <b>ó</b> +dad ~ pr- <b>ó</b> +dad
OBL	mr- <b>á</b> + <b>da-</b> ~ tr- <b>á</b> + <b>da-</b> ~ pr- <b>á</b> + <b>da-</b>	mr-e+dades- ~ tr-e+dades- ~ pr-e+dades-

*mrídaj* ‘my mother’ will be *mrída-*. This is, however, not the case in Vend Romani where the pronoun constituent takes the feminine oblique marker *-a*, which becomes lengthened analogically to the nominative forms, e.g., *mr-á da-tar* ‘my-OBL.F.SG mother-ABL; from my mother’, cf. *mr-a čha-tar* ‘my-OBL.F.SG daughter-ABL; from my daughter’. Note that the pronoun element of the compound is directly followed by the head noun *daj* ‘mother’ in cases<sup>5</sup> other than nominative, e.g., long vowel in *mrá + da-ha* ‘my.OBL.F.SG + mother.OBL.F.SG-INS; with my mother’, but short vowel in *mra moštóvna da-ha* ‘my.OBL.F.SG step-mother.OBL.F.SG-INS; with my step-mother’.

The development of long vowels in the monosyllabic possessive pronouns of Vend Romani is illustrated in table 4. First, the nominative nouns *daj* ‘mother’ and *dad* ‘father’ have become compounded with the pronouns *mr-* ‘my’, *tr-* ‘your’ and *pr-* ‘his/her own’. Second, the occurrence of long vowels has analogically extended to the feminine oblique forms of these pronouns.

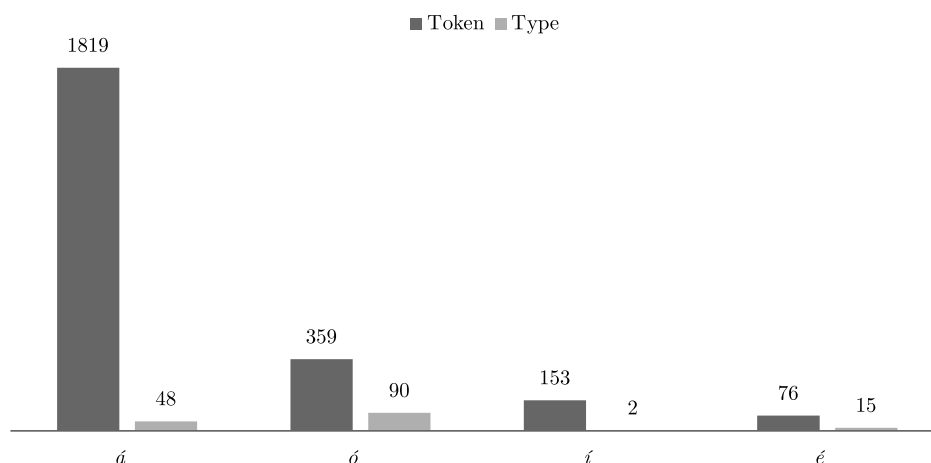
**Table 4:** Development of vowel length in possessive pronouns

	-MOTHER	-FATHER
	↓	↓
Compounding	my.NOM your.NOM his/her_own.NOM	my.NOM your.NOM his/her_own.NOM
Analogical extension	my.OBL your.OBL his/her_own.OBL	—

<sup>5</sup> The grammatical cases based on the oblique stem include the accusative (also called “independent oblique”, zero marked), genitive (*-ker-*, *-ger-*), dative (*-ke*, *-ge*), ablative (*-tar*, *-dar*), locative (*-te*, *-de*), and instrumental (*-ha*, *-ca*).

But what are the constraining factors of this change? In other words: Why has the vowel become long in the first person and second person singular pronouns as well as in the third person singular reflexive pronoun, but not in the masculine oblique forms of pronouns (e.g., *mre dades-*, not *mré dades-* ‘my father.OBL’), or in other possessive pronouns (e.g., *leskeri daj*, not *leskerí daj* ‘his mother’)? And why is the sound change limited to the nouns ‘mother’ and ‘father’ to the exclusion of nouns for other family members (e.g., *mri čhaj*, not *mrí čhaj* ‘my daughter’, or *mro phral*, not *mró phral* ‘my brother’)?

The described change in vowel length seems to be driven by certain phonological and semantic constraints: The former may account for the absence of a long vowel in the oblique masculine forms, e.g., *mr-e*, not *mr-é* ‘my.OBL.M.SG’. As shown in figure 1,<sup>6</sup> the sound *é* occurs significantly less frequently in word-final position than the sounds *á*, *ó*, *í*.



**Figure 1:** Token and type frequency of final *á*, *ó*, *í*, *é* in the Vend Romani variety of Kisbajom

Moreover, the final *é* appears exclusively in German and Hungarian-borrowed nouns, such as in *khafé* < German *Kaffee* ‘coffee’, or *tévé* < Hungarian *tévé* ‘television’. By contrast, the final *á*, *ó*, *í* sounds are additionally, or exclusively, encountered in inherited words, e.g., *adá* ‘this’, *asó*

<sup>6</sup> The data shown in figure 1 come from the Vend Romani variety spoken in Kisbajom. The Kisbajom Romani sample comprise 66 000 word-form tokens. The monosyllabic possessive pronouns are excluded from these data.



‘such’, *adí* ‘today’. Although the final *í* indicates the lowest type frequency,<sup>7</sup> its token frequency is double of that of *é* (figure 1). According to this, it is highly unlikely that the sound *é* would occupy the final position, especially in an inherited word, such as the masculine oblique possessive pronoun.

The number of syllables seems to also play a decisive role in the introduction of long vowels into possessive pronouns. It has been shown that only monosyllabic pronouns are influenced by the innovation, while those with more than one syllable have remained unchanged, e.g., *a.ma.ra da-*, not *a.ma.rá da-* ‘our mother.OBL’.

Finally, the spread of long vowels is semantically constrained by the lexical field comprising the kinship terms ‘mother’ and ‘father’. This may be due to a general tendency towards analogical change in the names denoting kinship. Winter (1969, 39ff), for instance, comparing several Indo-European languages, concludes that analogical change often takes place in the semantically closely related sets of kinship terms, such as mother : father. He provides an example of this type of change found in Tocharian, where the root vowel in *mācer*/*mācar* ‘mother’ is analogically taken over from *pācer*/*pācar* ‘father’. Winter cites another example from Old Church Slavonic, where the accusative form of ‘mother’ has been replaced by the genitive form in agreement with the inflectional pattern of ‘father’, i.e., *otica* ‘father.GEN/ACC’  $\Rightarrow$  *matere* ‘mother.GEN/ACC’ < \*‘mother.ACC’. In South Central Romani, we find similar analogical change in the kinship pair brother : sister. For instance, instead of the inherited terms *phral* ‘brother’ and *phen* ‘sister’, the borrowed *ečč-o* (< Hung. *öcs* ‘younger brother’) ‘brother’ and its feminine derivation *eč-kiňa* ‘sister’ is used in Dunajská Streda Romani (Slovakia), *táštvír-o* (< Hung. *testvér* ‘sibling’) ‘brother’ and its feminine derivation *táštvír-kiňa* ‘sister’ in Vlčany Romani (Slovakia), or *endán-i* (< *endáňi* ‘kin’) ‘brother’ and its feminine derivation *endán-kiňa* ‘sister’ in Mátraverebély Romani (Hungary). The first example is particularly interesting because the female counterpart is derived from the Hungarian-origin stem *eč-* with the original meaning of ‘younger brother’, instead of adapting the respective Hungarian terms *húg* ‘younger sister’ or *nővér* ‘older sister’. The strategy to derive the female counterpart from the male term of the kinship relation guarantees the formal similarity of these two terms. Thus, the formal or structural similarity of the kinship terms developed through analogy appears to be rather common, at least in the Indo-European languages.

<sup>7</sup> The sound *í* in final position is only attested in the inherited (contracted) nouns *dí* ‘day’ and *adí* ‘today’.

#### 4. Analogical extension of vowel length in the imperfective suffix

One of the main isoglosses that distinguish the South and North Central dialect groups of Romani is the imperfective suffix, which generally has the form *-ahi* in the former and *-as* in the latter (e.g., Elšík et al. 1999, 351). This suffix is used to form the imperfect, as well as the formally identical potential conditional, and the irrealis conditional. In most varieties of South Central Romani the suffix *-ahi* is realized with a short vowel, e.g., *garuv-es-ahi* ‘hide-2SG-IMPF; you were hiding; you would hide’ and *garudʔ-al-ahi* ‘hide.PFV-2SG-IMPF; you would have hid’ in Nógrádszakál Romani (Hungary). In some other varieties, the first vowel of the suffix is either long (*-áhi* or *-áj* in the Žitný ostrov region, Slovakia), or the short and long variants are in free variation (*-ahi* ~ *-áhi* in Pest county, Hungary). In most Vend Romani varieties, as well as in a few South Central varieties of Austria and Slovakia, the long and short vowel forms of the suffix have become functionally distinct. In these varieties, the suffix *-ahi* is reserved for the imperfect/conditional (e.g., *kereh-ahi* ‘you were doing; you would do’), while the marker *-áhi* is used to form irrealis conditional (e.g., *kerdʔal-áhi* ‘you would have done’). Compare the imperfective and irrealis forms of the verb *már*-‘to beat’ in Vend Romani, in table 5:

**Table 5:** Inflectional markers of imperfect and irrealis

	Imperfect	Irrealis
1SG	már(-)áhi	márdʔ-um-áhi
2SG	már-eh-ahi	márdʔ-al-áhi
3SG	már-l-ahi	márdʔ(-)áhi ↕
1PL	már-ah-ahi	márdʔ-am-áhi
2/3PL	már-n-ahi	márdʔ-en-áhi

In Vend Romani, the imperfect is formed by attaching the personal concord markers (*-a(v)-*, *-eh-*, *-l-*, *-ah-*, *-n-*) as well as the remoteness suffix *-ahi* to the stem. The first-person form is irregular, as it results from the contraction of the personal concord marker *-av-* and the suffix *-ahi*, i.e., *már(-)áhi* < \**már-av-ahi* ‘I was beating’. The irrealis consists of the perfective form of the verb stem (e.g., *márd-*) followed by the personal concord markers (*-ʔum-*, *-ʔal-*, *-ʔa-*, *-ʔam-*, *-ʔen-*) and the long vowel counterpart of the imperfective suffix *-ahi*. In this paradigm, the third person singular form is irregular (see the framed form in table 5) due to the contraction of

the personal marker *-a* and the suffix *-ahi*, i.e., *márdʔ(-)áhi* < \**márdʔ-a-ahi* < \**márdʔ-ah-ahi*.

As shown in table 5, the stress does not interfere with vowel length of the remoteness suffix. Instead, the stress generally falls on the personal concord marker, or less commonly, on the preceding syllable due to the elision of *e* in the third person singular and the second/third person plural markers, e.g., *már-l-ahi* < \**már-el-ahi*. The first syllable of the remoteness suffix, since it has been merged with the personal marker, has become stressed in the contracted forms, that is, in the first person singular imperfect and the third person singular irrealis. We argue that the latter form was the trigger for the long vowel becoming generalized – or analogically extended – across the whole irrealis paradigm (table 6).

**Table 6:** Development of vowel length in the irrealis forms

	Person	Form
Contraction	3SG	*stem-ah-ahi > stem-áhi
	↓	↓
Analogical extension	2SG	*stem-al-ahi > stem-al-áhi
	2/3PL	*stem-en-ahi > stem-en-áhi
	1SG	*stem-um-ahi > stem-um-áhi
	1PL	*stem-am-ahi > stem-am-áhi

Following the markedness theory (see, e.g., Jakobson 1939; Croft 1990; Greenberg 1966), we can determine that the third person singular is the unmarked<sup>8</sup> member of the irrealis paradigm, because the morphological boundary between the personal marker *-a* and the suffix *-ahi* has become blurred. This is also supported by the results of Elšík & Matras (2006, 361–362) based on a sample of various Romani dialects that the singular and, to lesser extent, the third-person form can be considered unmarked or, according to their terminology, “default values” in Romani. Returning to our example, the length of the vowel has become the only indicator of the underlying personal marker in third person singular. In contrast, both the personal and the remoteness suffixes are easily identifiable in the other persons. Thus it appears that the third-person suffix *-áhi* has been reanalysed as an irrealis suffix and, subsequently, extended

<sup>8</sup> The term *unmarked* refers here to the shortest, “least marked” or zero-coded elements of the respective paradigm that occur more frequently than the marked forms.

through the paradigm. In other words, the historical remoteness marker *\*-ahi* has split into two distinct markers in Vend Romani: into imperfective *-ahi* and irrealis *-áhi*. As a consequence, the irrealis third person singular form has become even less marked than before, since the personal marker ceased to be encoded by the long vowel. This form may therefore be analysed as consisting of a perfective stem and an irrealis suffix (in long vowel), while the person is zero-coded, e.g., *márd'-áhi* 'beat.PFV-0-IRR' < *\*márd'(-)áhi* 'beat.PFV-3SG.IRR'.

The question arises, therefore, as to why the contracted first person singular form has not triggered similar vowel lengthening in the imperfective paradigm, though it is the unmarked member in the respective paradigm. We assume that it is more essential to maintain the distinction between the imperfective and irrealis paradigms than the restriction imposed by the person/number value.

## 5. Summary

Vowel length is not a phonological feature that is common to Romani in general, since it only exists in the Romani dialects that have been in contact with languages exhibiting vowel length. This is the case in Vend Romani, where distinctive vowel length has developed as a result of contact with Hungarian. As a consequence, vowel length has become phonemic in Vend Romani.

This paper has described how the vowel length distinction, after its reintroduction into Vend Romani, may have spread due to the process of analogical extension. In Vend Romani, this process accounts for the development of long vowels in the monosyllabic possessive pronouns in combination with the words for 'mother' and 'father', as well as the vowel length distinction in the remoteness marker. The latter case is especially interesting in that the vowel of the remoteness suffix has become long throughout the irrealis paradigm (i.e., *-áhi*), while the suffix with short vowel (i.e., *-ahi*) has been reserved exclusively for the imperfect. This means that vowel length has become functionally relevant, and has therefore enriched the grammatical structure of Vend Romani.

It has been shown in the case of possessive pronouns that the spread of long vowels is limited by phonological and semantic constraints. Since long *e* is not permissible in word final position in inherited Romani words, the masculine oblique forms of possessive pronouns (in final *e*) do not lengthen. Furthermore, the change has affected the monosyllabic pronouns, but not the polysyllabic pronouns. Finally, the fact that the word-final

vowel lengthens only in combination with the closely related kinship terms for ‘mother’ and ‘father’ points to a semantic constraint.

As for the case of the remoteness marker, we find long vowel forms in both the irrealis and imperfective paradigms, which have emerged through contraction. While the third-person long vowel form in the irrealis has spread throughout the whole irrealis paradigm, the first-person long vowel form of the imperfect has not triggered a similar change in the respective paradigm. We assume that the differentiation between the imperfective and irrealis paradigms has prevented the spread of the long vowel form to the imperfective forms.

### Acknowledgements

We wish to thank Michael Beníšek for his valuable suggestions on the draft version of the paper.

### References

- Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In B. D. Joseph and R. D. Janda (eds.) *The handbook of historical linguistics*. Malden, MA & Oxford: Blackwell. 425–440.
- Blevins, James P. and Juliette Blevins. 2009. Introduction: Analogy in grammar. In J. P. Blevins and J. Blevins (eds.) *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press. 1–12.
- Bodnárová, Zuzana. 2013. Overview of Vend Romani in Hungary. In B. Schrammel-Leber and B. Tiefenbacher (eds.) *Romani V. Papers from the Annual Meeting of the Gypsy Lore Society, Graz 2011*. Graz: Grazer Romani Publikationen (GRP02). 29–4.
- Boretzky, Norbert. 1998. Der Romani-Wortschatz in den Romani-Misch-Dialekten (Pararomani). In Y. Matras (ed.) *The Romani element in non-standard speech*. Wiesbaden: Harrassowitz. 97–132.
- Boretzky, Norbert and Birgit Igla. 1993. Lautwandel und Natürlichkeit. Kontaktbedingter und endogener Wandel im Romani. Arbeitspapiere des Projektes “Prinzipien des Sprachwandels” 15. Essen: Universität GH Essen, Fachbereich Sprach- und Literaturwissenschaften.
- Boretzky, Norbert and Birgit Igla. 2004. *Kommentierter Dialektatlas des Romani*. Wiesbaden: Harrassowitz.
- Croft, William. 1990. *Typology and universals*. Cambridge: Cambridge University Press.
- Elšík, 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 in Österreich im interdisziplinären und internationalen Kontext*. Klagenfurt: Drava. 277–390.

- Elšák, Viktor and Yaron Matras. 2006. *Markedness and language change: The Romani sample*. Berlin & New York: Mouton de Gruyter.
- Greenberg, Joseph H. 1966. *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.
- Hock, Hans Heinrich. 1986. *Principles of historical linguistics*. Berlin & New York: Mouton de Gruyter.
- Jakobson, Roman. 1939. *Signe zéro*. *Mélanges de linguistique offerts à Charles Bally*. Geneva: Librairie de l'Université.
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Miklosich, Franz. 1872–1880. *Über die Mundarten und Wanderungen der Zigeuner Europas X–XII*. Wien: Karl Gerold's Sohn.
- Rézműves, Melinda. 2006. *O éorro rrom taj o borzo: Rromane paramíci – Cigány mesék* [The poor Gypsy and the hedgehog: Romani tales]. Budapest: Fővárosi Cigány Önkormányzat.
- Sowa, Rudolf von. 1887. *Die Mundart der Slovakischen Zigeuner*. Göttingen: Vandenhoeck und Ruprecht's Verlag.
- Vekerdi, József. 1984. The Vend Gypsy dialect in Hungary. *Acta Linguistica Academiae Scientiarum Hungaricae* 34. 65–86.
- Winter, Werner. 1969. Analogischer Sprachwandel und semantische Struktur. *Folia Linguistica (Acta Societatis Linguistica Europaeae)* 3. 29–45.