

Mikhail Oslon*

Reconstruction of Proto-Romani Demonstratives

Summary. Romani demonstratives (‘this’, ‘that’) show great diversity across dialects. The number of demonstrative pronouns in some living dialects is quite impressive (e.g., at least 17 in Russian Romani), and in terms of type of deixis they represent (in some dialects) a four-term system: ‘this [visible]’, ‘that [imaginary]’, ‘this [of two or more, visible]’, ‘that [of two or more, imaginary]’, which has been convincingly shown by Y. Matras, who has also proposed a reconstruction of their proto-forms. His reconstruction is now generally accepted, although it violates, in many respects, the comparative historical procedure. The present article gives an overview of the systems of demonstrative pronouns across Romani dialects and proposes a new reconstruction, which yields two sets of Proto-Romani demonstratives, inflected and uninflected, and suggests that all modern systems are reducible to a single Proto-Romani state, the number of innovations in individual dialects being relatively small.

Keywords: Romani, deixis, demonstratives, pronouns, diachrony, Old Indic, Middle Indic, reconstruction

1. Internal reconstruction

Romani has a complex and diverse system of deictic words, whose original (Proto-Romani) state seems, at first glance, to elude reconstruction. However, upon closer examination, deictics turn out to be one of the most regular, conservative, and uniform fragments of Romani morphology (we will only focus on demonstrative pronouns meaning ‘this’, ‘that’, etc.).

1.1. Morphological build-up

R: root

All demonstratives contain one of the two “primitive” deictic particles: *a-* or *-o-*, which we will call the “root”. They are often opposed to each other in terms of “source of knowledge”: *situational* vs. *discourse* (Matras 1994:

* I thank Michael Beníšek, Viktor Elšík, Kirill Kozhanov, Svetozar Lashin, Yanush Panchenko, Natalia Pimenova, Julieta Rotaru, and Galina Yavorska for discussion and corrections. Special symbols: * reconstructed or unattested form; ** non-existent (in some cases, expected) form; >/< phonetically yields/reflects; =>/<= becomes/replaces analogically; <=> analogically influenced by; ~ is in free variation with; : alternates with; | omission of repeated part to the left.

43–67; Matras 1998),¹ which in some dialects (secondarily) correlates with proximal vs. distal deixis.

C₁ prefix (*d-* or *k-*)

Consonants *d-* or *k-* prefixed to the root yield the opposition of “focus”: *non-discrete* (general deixis) vs. *discrete* (one of two or more items). Thus, a four-term system obtains:

<i>da</i> *‘this, visible’	<i>do</i> *‘that, in an account’
<i>ka</i> *‘this of two or more, visible’	<i>ko</i> *‘that of two or more, in an account’

This entire set is attested in at least one dialect, Finnish Romani, while the rest of the dialects contain fragments or none of it.

V₁ prefix (= R-)

Another vowel copying the root is prefixed, yielding the forms *ada*, *odo*, *aka*, *oko* (roughly, *‘this here’, etc.).

C₂ prefix (*k-*)

Another consonant *k-* (not *d-*!) is prefixed: *kada*, *kaka*, *kodo*, etc., a further deictic element.

V₂ prefix (= R-)

One more copy of the root: *akada*, **okodo* (only in the non-discrete series, i.e. there are no ***akaka*, ***okoko*).

Demonstratives formed by the above elements are indeclinable: they refer to both genders and numbers and do not agree with the noun they modify, cf. Russian Romani *da rom* ‘this Rom[ani man]’, *da romny* ‘this Romani woman’, *da romà* ‘these Roms’, *da romès-tyr* ‘from this Rom’ (ablative).

PS: personal suffix

This is a consonant suffixed to the root, marked for gender and number, in the direct case: *-v* (m.sg.), *-j* (f.sg.), *-l* (pl., both genders). Simple root + personal suffix combinations act as personal pronouns, cf. Crimean Romani *ov* ‘he’, *oj* ‘she’, *ol* (beside *on*; see § 4) ‘they’ (rarely *av*, etc.). The indirect forms contain, instead of *-v*, *-j*, *-l*, suppletive segments identical to the corresponding forms of the personal pronouns, e.g. Kald. dir. m.sg. *ko|(v)*, obl. *ko|lé(s)* (cf. *le(s)* ‘him’); f.sg. *ko|j*, obl. *ko|lá* (cf. *la* ‘her’).

E: extension (*-a*)

The rightmost affix attached to direct personal forms is a “deictic extension” *-a*, e.g. *kov|á*. It is not attached to oblique forms.

¹ This dichotomy is common in the related languages, e.g. in Shina (Schmidt, Kohistani 2008: 87).

Here is a sample of the general layout filled to the maximum extent:

		V ₂	C ₂	V ₁	C ₁	R	PS	E
m.sg.	dir.:	<i>a-</i>	<i>k-</i>	<i>a-</i>	<i>d-</i>	<i>a</i>	<i>-v</i>	<i>-á</i>
	obl:	<i>a-</i>	<i>k-</i>	<i>a-</i>	<i>d-</i>	<i>a</i>	<i>-lé(s)</i>	

Below, we summarize the available data from various Romani dialects in tabular form (see § 2). The tables are divided into four sections according to two dichotomies:

- focus: (1) non-discrete (simply ‘this’) and (2) discrete (‘this of two or more’);
- source: (1) situational (‘visible’) [some dial. ≈ near] and (2) discourse (‘imaginary’) [≈ far].

situational/ non-discrete	discourse/ non- discrete
situational/ discrete	discourse/ discrete

Each section has several columns (suffixes) and rows (prefixes):

- suffixes: (1) \emptyset ; (2) -PS; (3) -PS-*a*;
- prefixes: (1) \emptyset -; (2) C₁-; (3) V₁-C₁-; (4) *k*-V₁-C₁-; (5) V₂-*k*-V₁-C₁-.

Anticipating the conclusions drawn below, we give all the attested demonstratives that fit into the structure outlined above. So far, this is not the result of an internal reconstruction, but rather a simple structural breakdown of the available forms (pronouns attested in at least two “greater” dialect groups, i.e. North, Balkan, Vlax, and Central, are in boldface). This will be our **initial schema**:

	\emptyset	-PS	-PS- <i>a</i>	\emptyset	-PS	-PS- <i>a</i>
\emptyset -	<i>a</i>	<i>a v j l</i>	<i>a va ja la</i>	<i>o</i>	<i>o v j l</i>	<i>o va ja la</i>
<i>d</i> -	<i>da</i>		<i>da va ja la</i>	<i>do</i>		<i>do va ja la</i>
<i>_d</i> -	<i>ada</i>	<i>ada v j l</i>	<i>ada va ja la</i>	<i>odo</i>	<i>odo v j l</i>	<i>odo va ja la</i>
<i>k_d</i> -	<i>kada</i>	<i>kada v j l</i>	<i>kada va ja la</i>	<i>kodo</i>	<i>kodo v j l</i>	<i>kodo va ja la</i>
<i>_k_d</i> -	<i>akada</i>		<i>akada va ja la</i>	<i>okodo</i>		<i>okodo va ja la</i>
<i>k</i> -	<i>ka</i>	<i>ka v j l</i>	<i>ka va ja la</i>	<i>ko</i>	<i>ko v j l</i>	<i>ko va ja la</i>
<i>_k</i> -	<i>aka</i>	<i>aka v j l</i>	<i>aka va ja la</i>	<i>oko</i>	<i>oko v j l</i>	<i>oko va ja la</i>
<i>k_k</i> -	<i>*kaka</i>		<i>kaka va ja la</i>	<i>*koko</i>		<i>koko va ja la</i>
<i>_k_k</i> -						

This yields a total of 42 items (including personal pronouns), which seems unusually high. However, no modern dialect has all of them (the most

is 17, e.g. in Russian Romani). Some of these often appear in a somewhat different form. However, as will be shown, internal reconstruction reveals a secondary nature of these discrepancies, most of which are thus eliminated.

1.2. Early solutions

Miklosich seems to have been the first to express the idea that declinable demonstratives, such as *kadó*, *kadává*, combine an “adverbial” part and a pronominal one: “... *kadó* ist die Verbindung eines Adverbs *kada* mit dem Pronomen *o*, *e*, im acc. m. *les* u.s.w. griech. *kadává* m. *kadajá* f. *kadalá*, *kadalé* pl.” (1872–80: V, 25).

Similarly, Sampson: “The demonstrative pronouns proper, *akava* ‘hic’, *odova* ‘iste’, *okova* ‘ille’ [...] are based upon the adverbial stems *ak-*, *od-*, *ok-*, conjoined in the nominative with the stem *-ov* (*-av*) m., *-oi* (*-ai*) f., *-ol* (*-al*) pl. of the 3rd personal pronoun” (1926: 163).

Lucid as this insight is, it leaves the issue of “long” (extended) vs. “short” forms unexplained.

1.3. Yaron Matras’s reconstruction

Matras’s approach to the issue consists in the derivation of “short” forms from the “long” ones: e.g., *ka* is seen as a reduction of *kava*, etc. (Matras 2000: 110–2; 2002: 103–12).

Matras postulates the proto-forms **ata* (m.), **ati* (f.), **ate* (pl.) (situational) and **ota*, **oti*, **ote* (discourse). These forms (we take the situational set) phonetically evolved into **alo*, **ali*, **ale* (**VtV > VIV* is a regular pre-Romani sound change), whereupon a deictic particle **(h)a* was suffixed to the direct forms, which, according to Matras, yields the following developments (Matras 2000: 112; 2002: 106–7):

m.sg.	<i>*ata</i> > <i>*alo</i> ;	<i>*alo-a</i> > <i>*alova</i>	> <i>*ava</i>
f.sg.	<i>*ati</i> > <i>*ali</i> ;	<i>*ali-a</i> > <i>*alija</i>	> <i>*aja</i>
pl.	<i>*ate</i> > <i>*ale</i> ;	<i>*ale-a</i>	> <i>*ala</i>

The oblique forms (with no **(h)a*) evolve as follows:

m.sg.	<i>*atas</i>	> <i>*ales</i>
f.sg.	<i>*ata</i>	> <i>*ala</i>
pl.	<i>*atan</i>	> <i>*alen</i>

The parallel discourse set was: **ova* (obl. **oles*), **oja* (obl. **ola*), **ola* (obl. **olen*).

The situational set, according to Matras, is preserved in South Balkan dialects (*av-dives* ‘today’ and *aj-rat* ‘tonight’, with the final vowel dropped, but

not necessarily, cf. Zargari *ava-berš* (Matras 2002: 108), while the discourse set is ubiquitous in personal pronouns (*ov* ‘on’, *oj* ‘she’, **ol* > *on* ‘they’, with the final vowel dropped) and in the article (*o*, *i*, *ol*) (Matras 2002: 112).

Later, “place deictics” *d-* and *k-* were prefixed, which gave such forms as: *d-ava*, *d-aja*, *d-ala* and *k-ava*, *k-aja*, *k-ala* (similarly: *d-ova*, etc.). These were then partially reduplicated (omitting *d-/k-* and the final vowel, i.e. particle), which yielded **av-d-ava*, **aj-d-aja*, **al-d-ala* (similarly: *ov-d-ova*, etc.). These were then simplified into *adava*, *adaja*, *adala* (and *odova*, etc.). This is the first version of Matras’s reconstruction (2000: 111). In the second version, instead of these pronouns, he uses adverbial pronouns *adaj*, *akaj* ‘here’ and *odoj*, *okoj* ‘there’, so now **adaj-ava* > *adava*, etc. (2002: 109).

This reconstruction has been accepted by Romani scholars and reproduced in a number of works (e.g. Granqvist 2011: 343). Yet, despite its ingenuity, it contains a number of errors. Its main flaw is that it is at variance with Romani historical phonology.

The starting dir.sg. forms (**ata*, **ati*) are anachronistic: to get **alo*, **ali*, one must start with OIA **ātaka-*, f. **ātikā²*, given that, e.g., Romani *-o* (corresponding to Hindi/Urdu *-ā*) invariably comes from OIA **-aka-* (cf. OIA *maṇḍaka-* > Romani *maṇṛo* ‘bread’)³. Matras has two identical proto-forms **ata* evolve simultaneously into m.sg.dir. **alo* and f.sg.obl. **ala⁴*.

But the real problems begin from the second column on. The change **ali²ja* > **aja* (via **-l-?*) is arbitrary and hard to imagine anywhere outside the North (it certainly cannot be Proto-Romani). Neither **alova* > *ava* nor **ali²ja* > *aja* are regular pre-Romani changes. As for the dropping of *-a* (to get *av*, *aj*, etc.), it, too, seems to lack grounds, at least at the “Common Romani” level, especially in view of the ubiquitous preservation of demonstratives in *-a* (co-existing with those without it). The change **aldala* > *adala* is irregular as well.

The second version of the reconstruction is better in that the contraction **adajava* > *adava* is indeed phonetic (cf. **daja* > *da*, obl.sg. of **daji* ‘mother’, but **odojova* > *odova* would have to be analogical). However, it is marred by the fact that the prefixing of *akaj*, *odoj*, etc. now requires apheresis of *a-*, *o-* to explain *dava*, *kova*.

Given that the only regular change in the entire reconstruction is the trivial **VtV* > *VIV*, the reconstruction lacks methodological rigour. We are forced to conclude that Matras’s reconstruction violates the comparative-historical procedure and must be rejected. But even from a broader perspective, his attempt to reduce a paradigm with suppletivism (e.g. **ava*, obl. **ales*) to a

2 No such forms are attested in OIA/MIA.

3 Matras repeatedly ignores this in his work (e.g. he has OIA *kala-* > Romani *kalo* ‘black’, Matras 2002: 39).

4 Perhaps what is meant is **āta* vs. **ātā* (vowel length added), so this is not critical.

regular one using patently irregular phonetic changes does not seem the most natural solution.

1.4. The proposed analysis

The internal structure of Romani demonstratives is apparent from their very synchronic build-up. The prefix-root combination is an indeclinable (uninflected) demonstrative pronoun, and the personal suffix is an inflected demonstrative pronoun (with suppletive direct vs. oblique forms), optionally extended only in the direct forms. Here is such a typical paradigm (Crimean; the root is in boldface, the personal suffix is separated by a vertical bar, the extension, by a hyphen); we compare these with the 3rd person pronouns and then with the article:

‘this’:	m.sg.	f.sg.	pl.
dir.	<i>ada v-á</i>	<i>ada j-á</i>	<i>ada l-á</i>
obl.	<i>ada lé(s)</i>	<i>ada lá</i>	<i>ada lé(n)</i>

Personal pronoun (Crimean):

dir.	<i>o v</i> ‘he’	<i>o j</i> ‘she’	<i>o l</i> ‘they’	(~ <i>on</i> ; only <i>on(e)</i>)
obl.	<i>le(s)</i>	<i>la</i>	<i>le(n)</i>	outside the Balkan group)

Definite article (Kalderaš, simplified):

dir.	<i>o</i>	<i>i ~ e</i>	<i>le ~ el</i>
obl.	<i>le</i>	<i>la</i>	<i>le</i>

Note that the personal suffix of both the demonstrative and personal pronouns is clearly similar to the article (for their common origin see § 4). In addition, we see that the homonymous forms f.obl. *adalá* vs. pl.dir. *adalá* differ in internal structure: only the latter has the *-a* extension, while in the former the *-a* is part of the personal suffix.

The impression of an extreme number of demonstratives can be mitigated by the treatment of these forms as more or less stable (but still free) combinations of words of the type English *this (right) here, that (over) there* etc., some of which have undergone univerbation (i.e. ended up having one stress only) in Proto-Romani. We find a typologically comparable situation in Romanian with 15 demonstratives (including personal pronouns), some with variants (only the m.sg. forms are given; variants are separated by “=”):

	<i>*illu</i> _+ <i>*hāc</i> (?)	<i>*istu</i>	_+ <i>*hāc</i> (?)
_>	<i>āl = el</i>	<i>āl-a</i>	<i>ǎst = ist = iest</i> <i>ǎst-a = ist-a = iest-a</i>
<i>*hāc</i> +>	(?) <i>al</i>	<i>aist = aiest</i>	<i>aist-a = aiest-a</i>
<i>*ecce</i> +>	<i>cel</i>	<i>cist</i> (?) = <i>cest</i>	<i>cest-a</i>
<i>*hāc</i> + <i>*ecce</i> +>	<i>acel</i>	<i>acel-a</i>	<i>acest</i> <i>acest-a</i>

Forms with the *-a* extension (a late addition, maybe from **hāc*) have internal inflection, cf. m.sg.dir. *acést-a* (< **hāc* + **ecce* + **istu* + **hāc*), f.sg.dir. *aceást-a* (< **hāc* + **ecce* + **ista* + **hāc*), m.sg.obl. *acéstui-a* (< **hāc* + **ecce* + **istūi* + **hāc*), etc. This system is similar to the Romani one, albeit with fewer cells. There are two roots (**ill-* and **ist-*; Romani has two as well), three possible prefixes (*a-*, *c-*, *ac-*; Romani has seven), a case ending (zero, *-ui*, *-ei*, etc.), and an extension (*-a*, also in indirect forms; in Romani only in direct forms).

Typologically, one can also draw a parallel, e.g., with Ukrainian: also two roots (*s-* and *t-* in *s-ej* ‘this’, now archaic, and *t-oj* ‘that’), several prefixes (e.g. *o-toj* ‘that there’, *on-toj* ‘that yonder’, *ot-sej* > *ocěj* ‘this here’, now usually *cej* ‘this’, etc.) and suffixes, still felt as separate particles (*toj ho* ‘that yonder’, *cej vo* ‘this here’, etc.). As a result, we get a fairly large number of combinations.

A very intriguing feature of Romani is the presence of a special expletive (hesitation marker) *kova* (discourse; discrete; personal-suffixed and extended) ‘thingamajig; whats’isface’. It is in this function that it is attested in many dialects, thus clearly being Proto-Romani.

2. Survey of the dialects

In the tables below, personal pronouns are underlined, as are segments deviating from the initial schema; the expletive is marked with a superscript “^E” or “^(E)” (when also functioning as a normal demonstrative).

2.1. North-Eastern

In the NE dialects (Russian, Polish, Lithuanian, Lotfitka = Latvian Romani), the non-discrete series is densely filled, while the discrete one is empty save for one or two cells. The three columns are filled unequally (the 2nd one only includes personal pronouns).

Russian Romani (Russia, Ukraine, Kazakhstan; Kozhanov forthc.-1):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
∅-				<i>jǒ v j nɛ́</i>		
<i>d-</i>	<i>da</i>		<i>da vǎ jǎ lǎ</i>	<i>do</i>		<i>do vǎ jǎ lǎ</i>
<i>_d-</i>	<i>adǎ</i>		<i>ada vǎ jǎ lǎ</i>	<i>odǒ</i>		<i>odo vǎ jǎ lǎ</i>
<i>k_d-</i>	<i>kadǎ</i>		<i>kada vǎ jǎ lǎ</i>	<i>kodǒ</i>		<i>kodo vǎ jǎ lǎ</i>
<i>_k_d-</i>	<i>akadǎ</i>		<i>akada vǎ jǎ lǎ</i>			
<i>k-</i>	<i>ka</i>					^(E) <i>ko vǎ jǎ lǎ</i>
<i>_k-</i>						
<i>k_k-</i>						

Russian Romani *ka* is used in admiring utterances. **Lotfitka** (Latvia; Mānušs et al. 1997: 337) is almost the same (but has, e.g., no *akadǎ*); it lacks *ka*, but has *ko* (see Mānušs et al. 1997: 73). **Polish Romani**: *dava, dova* (Matras 1999: 11), as well as (indecl.) *da* (*da gil'a* ‘these songs’, RMS: PL-003). **Lithuanian Romani**: *dava, adava, dova, odova* (Tenser 2005: 19) and (indecl.) *da, ada, do, odo*; (decl.) *kadava, kodova* (Tenser 2008: 94).

Indeclinable deictics tend to be viewed as “simplified” from the inflected ones: “In all of the NE dialects there is an option of simplifying the inflected forms of demonstratives to (*a*)*da* and (*o*)*do*, which are not inflected for gender, number or case” (Tenser 2008: 94). Such a “simplification” is hard to imagine, since, e.g., for the replacement of obl. *dalés* by *da*, one would have to posit a non-phonetic “apocope” of the entire morphologically relevant part *-les*, which seems quite unnatural. Whatever the case, indeclinable deictics exist well beyond NE dialects (see below). For “Proto-NE Romani” we should reconstruct the Russian Romani system plus *ko* (see § 3).

2.2. North-Western**Finnish Romani** (“Kaale”, Finland; Granqvist 2009: 106):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–				<i>jo u i n</i>		
<i>d-</i>	<i>da</i>		<i>da vva la</i>	<i>do</i>		<i>do vva la</i>
<i>_d-</i>	<i>ada</i>		<i>ada vva la</i>	<i>ado</i>		<i>ado vva la</i>
<i>k_d-</i>						
<i>_k_d-</i>						
<i>k-</i>	<i>ka</i>			<i>ko</i>		<i>ko vva la</i>
<i>_k-</i>	<i>aka</i>			<i>ako</i>		
<i>k_k-</i>						

The leftmost vowel of the prefix here is always *a-*. If we take this to be a secondary generalization, then, e.g., *ako* \Leftarrow **oko* (\Leftarrow *aka*). Indeclinability can hardly be considered a local innovation, since (as in NE) there are no source models for it in the dominant (Finnish) language, which in this case has had another effect – the loss of gender distinctions in extended forms: sg.m. = f. *davva*. **Sinti** (only decl.): *dáva*, *dóva*, *káva*, *kóva* (Germany; Finck 1903: 31)⁵; the same plus “short” forms *ka|u|j|l*, *ko|j|l*⁶ (Germany; Holzinger 1993: 74); **Manush**: only *k|āvá|ajá|alá* (with length alternation), *kōvá*, along with “short” *ka|j|l*, *ko|j|l* (France; Calvet, Delvoye, Labalette 1970: 73). However, Sinti appears to have indeclinable *da* and *ko* (RMS: RO-022):

572: *dikhjom da manušes ti u vāverdis*

‘I saw the same man the next day as well’

768: *majbraven ko džuklen fun māru gap*

‘We drive away the dogs from our village’

Welsh Romani (Sampson 1926: 164):

	-∅	-PS	-PS- <i>a</i>	-∅	-PS	-PS- <i>a</i>
–				<i>jo v i n</i>		
<i>d-</i>				<i>dā-</i>		<i>do vā īā lā</i>
<i>_d-</i>				<i>odā-</i>		<i>odo vā īā lā</i>
<i>k_d-</i>						
<i>_k_d-</i>						
<i>k-</i>			<i>ka vā īā lā</i> ⁷			^E <i>ko vā īā lā</i>
<i>_k-</i>	(<i>aka-</i>)		<i>aka vā īā lā</i>			<i>oko vā īā lā</i>
<i>k_k-</i>						

Here, the non-discrete situational section (***da(va)*, etc.) is empty, which is clearly due to a secondary elimination. Indeclinables are only bound: (*o*)*dā-kai*, *-kā* ‘he, she, they who, that which’. As we see, (*o*)*dā* has the root *-ā* instead of **-o* (cf. the original root, but secondary prefixes in Finnish Romani *ado*, *ako*, see above). The indeclinable *aka-* (~ *ake-*) is less clear (cf. (*a*)*kadives* ~ (*a*)*ke-*, *kid-*, \langle keḍ \rangle /*kəd-*/ ‘today’, (*a*)*karat*, *kerat* ‘tonight’).

Sampson gives \langle vāvā \rangle , \langle dovā \rangle as due to apheresis from *akavā*, *odovā*, which may or may not be correct, yet he considers ^E*kovā* quite separate from *okovā* (Sampson 1926: 158). A salient feature is the mobile stress depending

5 The same set is found in the isolated **Dolenjski** dialect (Slovenia; Cech 2006: 34).

6 The short forms are due to a recent apocope (so they don’t belong in our 2nd column), since an original **kov* would have yielded **kob* (as Sinti *job* ‘he’ in Holzinger, but still *jo*v in Finck).

7 Variants for f.sg.: *akaiā* ~ *akīā*.

on predicative vs. attributive position: <akáva mūrš> ‘this man’, but <akavá s’ō mūrš> ‘this is the man’ (we will return to this in 2.5.3).

For an abstract “Proto-NW Romani” we would then reconstruct an eight-item indeclinable set (*da, do, ada, *odo; ka, ko, aka, *oko*), as well as two rows of declinables in each section of the table (so *dava, adava*, but not **kadava, *akadava*). This may mean that these longer sequences have not been lexicalized in NW (yet they clearly are not a NE innovation, as they, at least *kadava*, are present in Vlax, see below).

2.3. South Balkan

Only here do we have *av*, f. *aj*, pl. *al*, which can partially serve as personal pronouns (like *ov, oj, ol ~ on*). The 2nd column contains non-extended personal forms that take a new (unclear) extension *-kha*.

Crimean (Ukraine, Russia; Торопов 2003; Toropov, personal communication):

	-∅	-PS	-PS- <i>a</i>	-∅	-PS	-PS- <i>a</i>
–		<i>a v- j- **</i>		<i>o v j l~n</i>		
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>	(<i>adà</i>)	<i>ada v- j- l-</i>	<i>ada và jà là</i>	(<i>odà</i>)		<i>odo va ja la</i>
<i>k-</i> <i>_k-</i> <i>k_k-</i>	(<i>akà</i>)	<i>aka v- j- l-</i>	<i>aka và jà là</i>	(<i>okà</i>)	<i>oko v- j- l-</i>	<i>ko va ja la</i> <i>oko va ja la</i>

The pronouns *av-* and *aj-* are found only in *av-d’ès* ‘today’, *aj-r’át* ‘tonight’. There are indeclinable substantivized *adà, odà* (← **odò?*), etc. ‘this, that’, which are also used attributively, but only in m.sg. (but we can assume this hasn’t always been the case). In the 2nd column (*av, ov*, with no *-a* extension): a new extension *-kha* has been added (both in dir. and obl.) *adav-khà, adaj-khà, adal-khà*, etc. (obl.sg./pl. *akal-khè*). **Ursari** (Romania; Miklosich 1872–80: XI, 18; probably once identical to Crimean): e.g. *adau-kha*, f. *adaj-kha*, etc., but also *kava*, pl. *kala* (KaHтpя 1970: 28), absent from Crimean. **West Bulgarian** (Minkov 1997: 77) has more forms in the 2nd column: *dav-kha, adav-kha, kav-kha, akav-ka*. Very similar is **Sofia Erli** (based on Gilliat-Smith’s texts; Boretzky 1998: 135), with an optional (?) contraction (?): *okoja* but *odi* (← **odoja/*odoj?*). So is **Ajios-Athanaios** (Greece; Sechidou 2011: 42), but it also has *kavà* and *odov-*.

“Roumelian” (Thrace; Paspatis 1870: 71):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–		<i>aj</i>	?		<i>o v j l</i>	?
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>			(<i>kada vǎ jǎ lǎ</i>)			<i>odo vǎ jǎ lǎ</i>
<i>k-</i> <i>_k-</i> <i>k_k-</i>		? <i>akav-</i>	<i>ak (av)ǎ jǎ lǎ</i>		? <i>okov-</i>	<i>ok(ov)ǎ jǎ lǎ</i>

The Rumelian system has undergone a series of analogical changes. The *-a*-extended forms *kadavǎ* are from the (Vlax-looking) Zapari dialect. The m.sg. forms *akavǎ* (not mentioned explicitly, but given in examples) and *okovǎ*, *odovǎ* are fully regular, but appear to have contracted variants, e.g. *akǎ*, with only the contracted f. *ak(h)jǎ*, pl. *aklǎ* (< ***akaja*, ***akale*?), as *okǎ*, etc.

The *-kha*-extended forms *akavk|ǎ |(h)ja |le*, etc. look like the result of the loss of internal inflection and transfer of the personal suffix to the extension, i.e., say: f. **akaj-kha* ⇒ *akav-khja*.

The rest of South Balkan dialects have no *kha*-extension.

Parakalamos (Greece; Matras 2003: 78):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–			<i>a va ja la</i>		<i>o v j l</i>	<i>o va ja la</i>
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>		<i>ada</i>				
<i>k-</i> <i>_k-</i> <i>k_k-</i>			<i>aka vǎ jǎ lǎ</i>			<i>oko vǎ ja la</i>

Remarkably, there probably is an indeclinable *ada*: “ER set **ada* apparently disappeared, but a trace of it is left in the expression *pe ada* ‘therefore’” (Matras 2003: 78). **Sepeçides** (Turkey; Cech and Heinschink 1999: 33) lacks *ada*, *ava*, *ova*, but has *adava*, *odova/oduva* (f. *odija*), *kava*, *kova/kuva* (f. *koja*). **Kosovo Arli** (North Macedonia, Kosovo: Boretzky 1996: 14): *adava*, *odova*, *akava*, ^(E)*kova*, *okova*. Note the phonetic variability in f.sg. *akaja/akeja*, *okoja/okeja/okija*.

Zargari (Iran; Baghbidi 2003: 133; Windfuhr 1970: 276):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–			<i>āvā aja ālā</i>			<i>ovā oja olā</i>
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>						
<i>k-</i> <i>_k-</i> <i>k_k-</i>	<i>ka</i>		<i>k āvā aja ālā</i> <i>āk āvā aja ālā</i>	<i>ko</i>		<i>k ovā oja olā</i> <i>ok ovā oja olā</i>

Only Windfuhr (1970: 276) mentions *ka* (obl. *kas*, secondary?) and *ko* (obl. *kos*).

2.4. North Balkan

Bugurdži (Kosovo; Boretzky 1993: 47):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–						<i>o v j n</i>
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>						
<i>k-</i> <i>_k-</i> <i>k_k-</i>			<i>ka vā ja la</i> <i>aka vā ja la</i> <i>kak ā ja ala</i>			<i>ko vā ja la</i> <i>oko vā ja la</i> <i>ko/uk ā ja ala</i>

Reduction: *kadā* < **kadva*?, cf. f. *kadja*, pl. *kadla*); or contraction: < **kadava* (cf. *phena* ‘I say’ < *phenava*, Boretzky 1993: 14).

2.5. North Vlax and the “North-Vlax umlaut”

In North Vlax, all demonstratives, except for personal pronouns, begin with *k-*, but relics of simpler forms can be seen, e.g., in Kald. *a-d’ès* ‘today’, *a-r’at* ‘tonight’, as well as in some other combinations (see 2.5.3).

Before going into any detail, a digression into Vlax historical phonetics is in order. A striking feature of Vlax is the diphthong *ej* (in both South and North Vlax) in place of *aj* in the rest of Romani, serving as the principal group-defining feature. Although it looks like a natural sound change, the evidence is more than scarce and comprises but two words: *dej* (other dial. *daj*)

‘mother’ and *čhej/šej* (other dial. *čhaj*) ‘daughter’. This phenomenon is usually interpreted as a lexically-conditioned (!) change *aj* > *ej*, given the existence of Vlach *raj* ‘lord’, *phabaj* ‘apple’, etc. with no such change (Boretzky 2003: 16).

However, if this is a sound change, we must be able to define it with no lexical restrictions: in both of the above words, an “umlaut” before *-i* is reconstructible: **daji* > Vlach *dej*; **čhaji* > Vlach *čhej*, as opposed to *raj* ‘lord’, etc., which never ended in **-i*.⁸ It does not depend on gender either, cf. Vlach *baj* f. ‘sleeve’ (not ***bej*). To be sure, two semantically related (and rhyming!) examples are insufficient to posit a sound change, but there probably are one or two more. The more conspicuous one is f.sg. *kadé(j)* ‘this’ (in some NV dialects, vs. *kadaj-* in the rest of Romani; see below for the corresponding m.sg. **kadav* and on phonetic difficulties).

In addition, there are forms in NV where traces of the umlaut are discernible indirectly. To identify them, we need to know that two vowel contractions took place in NV:

- (1) **eja* > *ēā* > *’ä* (probably regardless of stress)
- (2) **āya* > *o* (probably only with stress on the first of the two syllables)

The “1st NV contraction” must have taken place upon the integration of Proto-NV dialects into the Romanian phonological system (cf. the characteristic Romanian diphthong *ēā*). At least in part of NV, it merged with *ja* and, at some point, was simplified into *ä*, which survives (under certain conditions) until now in the Rakhiv dialect (Ukraine, see Ослон forthc.). Elsewhere, it merged with /a/ (palatalizing the preceding consonant)⁹ or (less often) with /e/ (as in Bukovina, see below)¹⁰.

The “2nd NV contraction” has nothing to do with Romanian and is fully reflected only in part of the NV: it is regular in verbs in some dialects, cf. Lovari *xos* ‘I ate’ < **xāyas*; *xo* ‘I will eat’ < **xāya*. It depends on stress, cf. Lovari, Rakhiv. *avāv* ‘I come’ (not ***ov*). In other dialects, including Kalderaš, this contraction does not appear in verbs (Kald. *xāvas*, *xāva*).

8 This would imply that “**daj*” and “**čhaj*” cannot be given as Proto-Romani (or “Common-Romani”) forms (as, e.g., in Matras 2004: 25). Hence, the practice of using living (but archaic-looking) forms as Proto/Common-Romani is, at least in these cases, to be abandoned (see ЭЦИЯ: s.v. **daji*; **čhaji*).

9 In Kalderaš (and some other, if not in all, NV dialects), this monophthongization was phonologically constrained: it only occurred after paired (“hard” vs. “soft”) consonants, e.g.: *dilēja* > *dil’ā* ‘fool!’ (rarely *dilēja*), but only *čorāja* ‘poor man!’, with variation under certain conditions: *phur’ā* ~ *phurāja* ‘old man!’ (Ослон 2018: 125).

10 Another example might be Kald. *s’a* ‘all’ (a variant of *sa*), which may represent **sej* (< **saji*) + **a* (this needs further examination).

2.5.1. Kalderaš-type dialects

Russian/Ukrainian Kalderaš (Ослон 2018: 250–2; Деметер and Деметер 1990):

	-∅	-PS	-PS- <i>a</i>	-∅	-PS	-PS- <i>a</i>
–	?? <i>a- ə- **</i>			<u><i>vò j n</i></u>		
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>	<i>ka dò d'à dalà</i>			<i>ko dò d'à dolà</i>		
<i>k-</i> <i>_k-</i> <i>k_k-</i>	<i>ka j l</i>			<i>ko j l</i> ** ** <i>ko l'à</i>		
	<i>ka kò t'à kalà</i>			<i>ku kò t'à kolà</i>		

The monosyllabic “endings” m.sg. *-o* and f.sg. *-’a* (unlike pl. *-ala/-ola*) require explanation. We can surmise that the “2nd Vlax contraction” (see above) has been eliminated in conjugation (so Kald. *xáva* ‘I will eat’, not ***xo*), but preserved in m.sg. *kadó*, *kakó* ‘this/that’. This requires the stress **kadàya*, **kakàya* (extended forms), cf. the expected *kadavà*, *kakavà* below. In turn, the f.sg. forms are the outcome of the “1st Vlax contraction”, so that *kad’á* < **kad’á* < **kadeá* < **kadeja* = **kadej* + **a* (extension). As noted above, *kadej* survives in some Kalderaš-type varieties, and we will show that it is due to “Vlax umlaut” (so we need an **...aj-i* protoform). Logically, we would expect a reflex of **kej* as well (here, we only have *kaj*) and, indeed, we will see it in the Bukovina dialect.

However, in this Kalderaš variety, there may be a trace of the umlaut in an even simpler form. If we assume that Kald. *ad’ès* ‘today’ < **avd’es* (cf. *garadó* < **garavdó* ‘hidden’), then it would exactly correspond to Crimean *avd’ès*. Kald. *ər’át* ‘tonight’ would then correspond to Crimean *ajr’át*¹¹, which would entail Kald. *ər’át* < **ej-r(’)at* with **ej-* < **aj-i*. The other (more likely?) possibility would of course be that *ə-* is the f.sg. article. The discourse forms *kodó*, f. *kod’á*, rhyming with *kadó*, *kad’á* must be secondary (**-oja* is not expected to contract).

Swedish Kalderaš (Gjerdman and Ljungberg 1963: 94) is much the same (although it lacks *ka*), but note the optional voicing in *kadó* ~ *gadó* and *kodó* ~ *godó* (yet only ^(E)*kova*). **French Kalderaš** (Calvet 2009) seems to lack pl. *kola*, but note *kaü* (2nd column). Phonetically, this behaves similarly to the combination of the “preposition” *ka* + article m.sg.dir. *o*. In part of Kalderaš, *o* after *-a* drops without a trace, cf. Moldovaja Kald. *ka řóm* ‘to the Rom’ (< *ka o řóm*; Ослон 2018: 199), but French Kald. *ka o řom* (Calvet 2009: 153).

11 The palatalization in Crimean *ajr’át* (vs. Crimean *rat* ‘night’) is unclear (from Vlax?).

Serbian Kalderaš (Boretzky 1994: 54):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–	<i>a-</i>			<i>vò (v) j n</i>		
<i>d-</i> <i>_d-</i> <i>k_d-</i>		? <i>kade(j) ?</i>	<i>kad avà ~ ò eja alà</i>	? <i>kode(j) ?</i>	<i>kod ovà ~ ò ojà olà</i>	
<i>_k_d-</i>						
<i>k-</i> <i>_k-</i> <i>k_k-</i>		? <i>ka j ?</i>	<i>ka và jà là</i>	<i>ko j l</i>	<i>ko và jà là</i>	
			<i>kak avà ~ ? ? ?</i>		<i>kuk ovà ~ ? ? ?</i>	

Free variation in f.sg. *kadeja* ~ *kadaja* (?), *kodoja* ~ *kodeja* ~ *kodja*; co-existence of *kakavà*, *kodovà* and *kadò*, *kodò*. The f.sg. form *kade(j)* has no *a*-extension.

Kišynjovarja (Russia, Ukraine; Кожанов forthc.-3):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–	<i>a-</i>			<i>ò v j n(è)</i>		
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>			<i>gada vâ ja lâ</i>			<i>godo vâ jà lâ</i>
<i>k-</i> <i>_k-</i> <i>k_k-</i>		<i>ka j l</i>		<i>ko j l</i>	<i>^Eko va ja la</i>	
			<i>kaka vâ jà lâ</i>			<i>kuko vâ jà lâ</i>

Only voiced: *gadavâ*, no ***kadava*. Note the stress *^Ekovâ*. No contracted forms of the type f.sg. *kad'â* (expected ***gag'â*), but rather only *gadajâ* (albeit *dej* ‘mother’, with umlaut). Very close to Kišynjovarja are the mutually closely related 19th-century dialects from the collection of tales and songs published by Constantinescu (1878, 2016; see also Miklosich 1872–80: XI, 18), which have, e.g. f.sg. *gadej*, *godej* (as well as *godòj*), also note *okouavèr* ‘the other’ (Rumanian *cel-(ă)l-alt*), pl.obl. *okolavré* (Constantinescu 1878: 85, 86).

2.5.2. Non-Kalderaš-type dialects

Czech/Slovak Lovari (Wagner 2012: 73):

	-∅	-PS	-PS- <i>a</i>	-∅	-PS	-PS- <i>a</i>
-	<i>a-</i>				<u><i>vou(j) (j) n</i></u>	
<i>d-</i> _ <i>d-</i> <i>k_d-</i> _ <i>k_d-</i>		** ** <i>kada j</i>	<i>kad o i ala</i>		** ** <i>kodo j</i>	<i>kod o i ola</i>
<i>k-</i> _ <i>k-</i> <i>k_k-</i>		<i>ka j j</i>			<i>ko j j</i>	
		** ** <i>kaka j</i>	<i>kak o i ala</i>		** ** <i>kuko j</i>	<i>kuk o i ola</i>

The PS pl. *-j* (instead of the expected *-l*), is unclear, e.g. *kada|j* (but *ka-dala* is present as well).

Polish Lovari (Pobożniak 1964: 49): only *kad|o |i |ala*; *kod|o |i |ola*; also *kuko*. Russian Lovari (Смирнова-Сеславинская and Цветков 2009): *kad|ò |j' |ól*; *kad|ò |j' |ól*. Whereas **áua > o* is entirely regular (whence *kado*, see above), f. *kadi*, *kodi*, *kuki* is probably due to an analogy with adjectives (cf. *bāro* ‘big’, f. *bāri*). The unextended pl. *kadol*, etc. may be from the 2nd column (details are unclear). **Rakhiv** (Ukraine, Ослон forth.): only *kad|ò*, |*á*, |*alá*; *kakò*; *kod|ò*, |*á*, |*alá*; *kukò*, as in Kalderaš.

Bukovina (then Romania; Miklosich 1872–80: V) has much the same forms as Lovari, but with the original ending in f. *kadè* (cf. with elision: *kad' aulín* ‘this palace’) < **kadā* (= Rakhiv) < **kadejā*. There is *ko*, f. *koj*, but no **ka(v)*, **kaj*. However, it is this latter form, but apparently with Vlach umlaut: **kej*, that can be recovered from combinations with *avér* ‘other’, namely (in Miklosich’s notation) “*téver*, *íar*”: “Vgl. die mir dunklen Redensarten *p' o tever lume*, *p' o íar lume* in die andere Welt” (Miklosich 1872–80: V, 7). This *te-* (with elision: *t'¹²*) is clearly the reflex of f. **kej*; hence, *téver* (for **-t'éver?¹³*) < **-kej-aver* can be the feminine counterpart of what would correspond to Kald./Lovari m. *káver* ‘other’ < **ko(y)-aver*, [now m. = f.], pl. *kol-avér* (calquing Romanian *cel(ă)lalt* ‘the other’, etc.)¹⁴. The incorrectly segmented “*p' o tever*” stands for **p[e] ot'éver* (since *lume* ‘world’ is feminine), i.e. *ot'éver* <

12 The form *íar* is contracted (Miklosich does not note the variant ***ar*, although cf. *avél ~ al* ‘comes; will be’).

13 Cf. in his next entry: “*tež*, *těž*, *těž* subst. m. Seide. *težésko*, *těžésté* seiden...” (Miklosich 1872–80: V, 59) < **k'ěž* ‘silk’ with and without palatalization.

14 The stress here (*káver*, f. **kéver*, but pl. *kolavér*) is not a problem, if we assume that it was determined by the “rhythmic law” after the contraction had been completed (**kóu-avér* > **kòavér* > *káver*) in Romania.

**ot'äver* < **okej'äver*. The problem is that we expect **okoj-äver* (inter alia, in view of the onset *o-*), but cf. Bukovina *kukò*, f. *ku'tè* (beside *kukè*), analogical to *kodé* (just as Kald. [*ku*]t'ä ⇐ [*ko*]d'ä). Hence, the form **(a)t'ä* < **(a)kejä* was lost, but the surviving analogical *ot'ä*- “materially” contains the segment **kej* (instead of *koj*).

2.5.3. North Vlux: discussion

The foregoing raises two main questions:

1. Why is it that the form f.sg. **kej* appears in combination with *avèr* ‘other’, but not on its own (only *kaj* is attested)?
2. How can contracted forms, e.g. *kadò*, coexist with, e.g. *kadavà*? Both are attested in Serbian Kalderaš (unless they are from different subdialects)?

The first question can be answered by assuming paradigm levelling. Our initial schema has the following unextended forms: *kav*, *kaj*, *kal*; *kov*, *koj*, *kol* (once again, this is not a reconstruction). In Vlux, we have no **kej*, but we do have Bukovina f. *-tèver* ‘the other’, which can be explained by the following scenario (here 1 = the “1st NV contraction”; 2 = centralization after velars, cf. Kald. **ker* > *kər* ‘do!’, see Oslon 2017; 3 = positional palatalization):

unextended: **kej* (2) > **kəj* (unattested)
 extended: **kej-a* (1) > **kēā* (3) > **k'ä* > **t'ä* (unattested)
 combined: **kej-aver*¹⁵ (1) > **kēāver* (3) > **k'äver* > **t'äver* > Buk. [*o*]tèver

These sound changes would have yielded: (unextended) *kav*, f. **kəj*, pl. *kal*; (extended) *kavà*, **t'ä*, *kalà*. The resulting morphological irregularity is then naturally eliminated by levelling:

	m.	f.		pl.
	<i>kav</i> (or [<i>></i>] <i>ka</i>)	<i>*kej</i> ¹⁶ > <i>*kəj</i> ⇒ <i>kaj</i>		<i>kal</i>
+ <i>-a</i> :	<i>kava</i>	<i>*keja</i> > <i>*t'ä</i> ⇒ <i>kaja</i>		<i>kala</i>

The phonetically regular **t'äver* was also eliminated in most dialects (but not in Bukovina), so, e.g., Kald. *káver* m. = f. (but pl. *kolavèr*!) (Деметер and Деметер 1990).

Note that f.sg. *kadèj* has been preserved (at least in Serbian Kalderaš, but also in a Russian Kalderaš variety: Jonešti *kadè(j)*, Oslon: fieldwork data), as well as in the extended Kald. *kad'ä*, etc. and *kat'ä*, which presupposes **kakej+a*. The corresponding m.sg. form is attested by Constantinescu: *kadau*.¹⁷

15 This is actually a replacement of **kojaver* (see above).

16 Miklosich (1872–80: XI, 18) erroneously gives “*kej*” from Constantinescu (1878: 42), where we find “*ke-i*”, “*kei*” in an Ursari song (a preposition + article combination, not a pronominal form).

17 In some dialects (but not in Kalderaš) this form, too, could probably give *kadò* (cf. *kamàv* ‘I want, I love’ > Buk. *kamò*, alongside *kamàu*, *kamàp*, Miklosich 1872–80: V, 26).

But how has this **kej/-dej* come about phonetically? For the umlaut to have occurred, there needs to have been a vocalic **-i* at the end (and not *-j*, cf. Vlax *baj* ‘sleeve’). This can be explained by an intermediate stage **ka-i/*da-i*, where *-i* is the f.sg. definite article. If so, non-Vlax *daj* vs. Vlax *-dej* ‘this (f.)’ is phonetically exactly parallel to non-Vlax *daj* vs. Vlax *dej* ‘mother’ (< **daji*).

As for the second question about the coexistence of *-ò* and *-avà*, an accentual explanation is possible. The development **-àua > -o* depended on stress (**-auà* did not contract). It is exactly in these demonstratives that we see a syntactically conditioned accentual variation in Welsh Romani (see Ослон 2014: 310 for details):

NV <i>kakavà</i>	< <i>*kaka<u>u</u>à</i>	cf. WelR <i>akavà</i> (predicatively)
NV <i>kakò</i>	< <i>*kaka<u>u</u>à</i>	cf. WelR <i>akàva</i> (attributively)

If so, in most NV dialects only one of these forms has survived. A problem remains: neither contracted nor umlauted forms are attested in Kišynjovarja (it only has *kadajà/gadajà* where we would expect **kadeja > **kag'a*). This may of course be due a levelling to the regular forms of the type *kodovà, kodojà* etc. (the opposite of what has happened, e.g., in Russian Kalderaš).

One must, however, acknowledge that the above points on Vlax umlaut are weakened by the presence of similar (but apparently much less regular) narrowings in Balkan, Central, and even in North-West (WelR) dialects: there, the expected **-daja, -doja* sometimes appear as **-dija, -dja, -di* (see tables).

2.6. South Vlax

Vlaxyccko (Ukraine, Russia; Панченко 2013: 10; Panchenko, personal communication):

	-∅	-PS	-PS- <i>a</i>	-∅	-PS	-PS- <i>a</i>
–				<i>(v)o v j n</i>		
<i>d-</i>						
<i>_d-</i>						
<i>k_d-</i>	<i>kadà</i>	<i>kad evà ejà elà</i>		<i>kodà</i>	<i>kod evà ejà elà</i>	
<i>_k_d-</i>	<i>akadà</i>	<i>akad evà ejà elà</i>		<i>akodà</i>	<i>akod evà ejà elà</i>	
<i>k-</i>		<i>** ka j **</i>	<i> ** ka jà là</i>		<i>ko j j</i>	<i> ** ko jà là</i>
<i>_k-</i>			<i> ** ** aka là</i>			<i> ** ** akò là</i>
<i>k_k-</i>						

This dialect (emigrated from Romania in the 18th century) has several in-declinables (partly with *-a* instead of **-o*) and, possibly, traces of Vlax umlaut: f. *kadejà*, secondarily also m. *kadevà*, etc. But **Servicko** (Ukraine, Russia,

outside Romania since the 17th century, Кожанов forthc.-d), has the expected *kadavà*, f. *kadejà*, pl. *kadalà*; *kodovà*, f. *kodejà*, pl. *kodolà*, displaying regular umlaut with only one analogy (**kodojà* ⇒ *kodejà*). This may reflect an older stage in Vlax.

Knjaževac Gurbet (Serbia; Ćirković and Mirić 2017):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–				<i>(v)o v j n</i>		
<i>d-</i> <i>_d-</i> <i>k_d-</i>	<i>gadà</i>		<i>gad avà ijà (ajà) alà</i>	<i>godà</i>	<i>godo và ijà (jà) là</i> (<i>god ò ì è</i>)	
<i>_k_d-</i>						
<i>k-</i> <i>_k-</i> <i>k_k-</i>	<i>k-/ga</i>		<i>k-/ga và jà là</i> <i>aka và jà là</i>	<i>go</i>	<i>go và jà là</i> <i>oko và jà là</i>	

Here, we see four indeclinables and somewhat ambiguous traces of Vlax umlaut. The paradigm *god|ò |ì |è* is unexpected, as *-o* “should” only be found in North Vlax, but this is certainly a secondary regularization to match adjective declension (cf. pl. *-e*). Similar sets are given for other **Gurbet** varieties (Kosovo; Leggio 2011: 78; Boretzky 1986: 202); **Agia-Varvara** (Greece; Iglu 1996: 39) has fewer forms, no indeclinables, but, in addition, it has *ada(v)à*, *kada(v)à*, *odo(v)à* (note also the variants f. *kajà* ~ *kaà* ~ *keà*, the latter with Vlax umlaut?); **Komotini** (Greece; Αλεξίου) is similar.

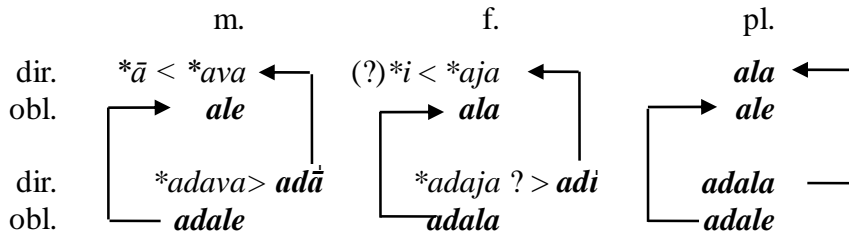
2.7. North Central

Central Romani deictics have undergone more analogical changes, so that, at first glance, they appear to violate our initial schema. Let us start with the best-described dialect, **Uzh** (Ukraine; Beníšek 2017: 234). Here are the full paradigms of all the 6 distinct demonstratives (dir.m.sg. forms in boldface):

		m.	f.	pl.
non-discrete:	dir.	<i>adǎ</i>	<i>adi</i>	<i>àla</i> ~ <i>adàla</i>
	obl.	<i>àle</i> ~ <i>adàle</i>	<i>àla</i> ~ <i>adàla</i>	<i>àle</i> ~ <i>adàle</i>
	dir.	<i>kadǎ</i>	<i>kadi</i>	<i>kàla</i> ~ <i>kadàla</i>
	obl.	<i>kàle</i> ~ <i>kadàle</i>	<i>kàla</i> ~ <i>kadàla</i>	<i>kàle</i> ~ <i>kadàle</i>
	dir.	<i>odǎ</i>	<i>odi</i>	<i>òla</i> ~ <i>odàla</i>
	obl.	<i>òle</i> ~ <i>odàle</i>	<i>òla</i> ~ <i>odàla</i>	<i>òle</i> ~ <i>odàle</i>
	dir.	<i>kodǎ</i>	<i>kodi</i>	<i>kola</i> ~ <i>kodàla</i>
	obl.	<i>kòle</i> ~ <i>kodàle</i>	<i>kòla</i> ~ <i>kodàla</i>	<i>kole</i> ~ <i>kodàle</i>

		m.	f.	pl.
discrete:	dir.	akā	<i>aki</i>	<i>akāla</i>
	obl.	<i>akāle</i>	<i>akāla</i>	<i>akāle</i>
	dir.	okā	<i>oki</i>	<i>okāla</i>
	obl.	<i>okāle</i>	<i>okāla</i>	<i>okāle</i>

As we see, the non-discrete series (*ad-*, *od-*, *kad-*, *kod-*) is represented by two forms in free variation in each slot, without and with *-da-*, except two (sg.dir.). The shorter forms like *á-le*, *á-l-a*, etc. look as if belonging to the paradigm of **a-v-a*, absent from this dialect. However, the stress and vowel length in, e.g., *adā* unequivocally point to an earlier **adava*. The stress retraction (common to all Central dialects, except, maybe, Plaščuna) took place prior to the contraction, so that **(ad)avá* > **(ad)áva* and then *adáva* > *adā*¹⁸. As for **ava*, which must have become **ā*, it was abandoned (being too short) in favor of *adā*. The same logic applies to f. **aja* and **adaja* > *adi* (the short end vowel is unclear, but stress points to contraction). Hence, in each of these four pronouns two distinct paradigms (e.g. **ava* and **adava*) have merged as a result of the elimination of **ā* and **i* (?) (living forms are in boldface):



The discourse/remote set must have then been restructured analogically: **odova* ⇒ *odā* (↔ *adā*), the root altered from **-o-* to *-a-*: **odola* ⇒ *odala*, etc., so that, functionally, it is now perceived as part of the “inflection”. The *kVdV*-series **kadava* > *kadā*, *kodova* ⇒ *kodā* is exactly parallel, except that, having merged with the originally discrete **kava*, **kova*, the discrete/non-discrete opposition was lost in this fragment of the system. The other discrete deictics **akava* > *akā*, **okova* ⇒ *okā* (↔ *akā*) have been fully preserved. Hence, the system, with the eliminated forms included (we only omit **(k)ova*, **(k)oja*), looks as follows:

18 Cf. also Uzh *keravas* > *kerās* ‘I was doing’ and *kerava* ~ *kerā* ‘I will do’ (with optional elimination of contraction).

Uzh (Ukraine; Beníšek 2017: 234):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–			*ā *i ala		<u>o v j n(e)</u>	** ** ola
d-			ad ā̇ i á la			od ā̇ i á la
_d-			kad ā̇ i á la			kod ā̇ i á la
k_d-						
_k_d-						
k-			*kā *ki kála			(*kā *ki) kó la
_k-			ak ā̇ i á la			ok ā̇ i á la
k_k-						

There are no traces of indeclinable pronouns in Uzh: even *adād'ive* ‘today’ contains *adā* < **adava* (secondarily?).

Deictics in **West Slovak** varieties have received several unequally reliable descriptions (Kalina 1882: 57; Sowa 1887: 70–2; Bourgeois 1911: 12; Lípa 1963: 92) analysed by Elšík, Hübschmannová and Šebková (1999). In some of them, there is the paradigm *ada*, f. *ada*, pl. *ala* (Kalina), and in others, *ada*, f. *aja*, pl. *ala* (Sowa) (same for *o-*). All of these systems are more or less reducible to the state illustrated by Uzh (albeit less transparent, due to lack of stress or length marks), with paradigms (partially) merging due to contractions. Moreover, an archaic state may have been captured by Kalina and Bourgeois, where contraction seems optional, cf. both *(ak)ada*, f. *(ak)ada* and *(ak)adava*, f. *(ak)adaja*¹⁹.

Note the existence of an indeclinable form: “at the time of v. Sowa, the nominative singular masculine short form could be used in the feminine, in the plural, or in the oblique; today, the short form *oda* is indeclinable in WSR [West Slovak Romani]” (Elšík, Hübschmannová and Šebková 1999: 344)²⁰. If this *oda* is old, it must be an alteration of **odo* (note the preserved ^E*kova*).

The very similar **Bergitka** (Poland; Rozwadowski 1936) has *ada*, f. *aja*, pl. *ála* (with initial stress and no vowel length), which may or may not point to an alternative scenario involving contamination rather than contraction. On the other hand, *okóva* (regular) and *okóda* (analogical) seem to be interchangeable (cf. only *akada* and no **akava*, but only ^E*kóva*).

19 An explanation requiring irregular changes: “The augment [= our “root”] vowel could be syncopated in some cases: *odoja* (or *odija*, or *odaja*) > **odja* > *od'a*, *odole* (or *odale*) > *odle*, and *odola* (or *odala*) > *odla*. After the syncope, the resulting consonant cluster could be simplified by dropping the *d*: **odja* > *oja*, *odle* > *ole*, *odla* > *ola*” (Elšík, Hübschmannová, Šebková 1999: 344).

20 Cf. also Kalina’s obscure note: “Avec les substantifs, ce pronom s’emploie pour tous le[s] cas du singulier *oda*, du plur. *ole*” (1882: 58).

Czech Romani (Ješina 1882: 23, 43):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–					<i>j o b j n</i>	.
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>			<i>ada va *ja *la</i>	(<i>odā?</i>)	<i>odov ? ?</i>	<i>do ba ja la</i> <i>odo va ja *la</i>
<i>k-</i> <i>_k-</i> <i>k_k-</i>			<i>ga ba ja la</i>			

The expected set *odova*, *odoja*, *odola* apparently has variants *odolo*, *odoli*, **odole* (parallel to the alternative personal pronoun set *lo* ‘he’, *li* ‘she’, *le* ‘they’, Ješina 1882: 21), which may be a secondary development, but cf. *olo*, *olo*, *oli*, **ole* (with no ***ova*, etc.). There are no indeclinables, but *oda*, f. *oda* may be a vestige of **odo*.

Plaščuna (Russia; Kozhanov forthc.-2):

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–					<i>j ò v j n(é)</i>	
<i>d-</i> <i>_d-</i> <i>k_d-</i> <i>_k_d-</i>	<i>adà</i>		<i>ada vâ jâ lâ</i>			<i>odo vâ jâ lâ</i>
<i>k-</i> <i>_k-</i> <i>k_k-</i>	<i>akadà</i>		<i>kada vâ jâ lâ</i>	<i>okodâ</i>		<i> ? ko jâ lâ</i>

This emigrant dialect (probably first brought to Ukraine in the 18th or 19th century) has indeclinable pronouns (but here one cannot exclude the influence of NE Romani). Unlike all other Central dialects, it shows no traces of the contraction **ava* > *ā* (cf. Plaščuna *kerava* ‘I will do’).

2.8. South Central

Burgenland (Austria; Halwachs and Wogg 2002: 17) only has *ada*, *oda*, and *ka*:

	m.	f.	pl.
dir.	<i>adà</i>	<i>aja</i>	<i>adala</i>
obl.	<i>adale</i>	<i>adal a ~ e</i>	<i>adale</i>
dir.	<i>ka</i>	<i>kija</i>	<i>kole</i>
obl.	<i>kole</i>	<i>kol a ~ e</i>	<i>kole</i>

The stress *adà* points to the contraction scenario (see also Halwachs, Wogg 2002: 6). The remote/discourse deictic *ka* clearly replaces **kova* (← **kava*), while f. *kija* is phonetically unclear (as is this form in most Central dialects).

Prekmurje (Slovenia; Baranja 2013: 45) has *adau* (< **adā*), *udau* (< **odā*), *ukau* (< **okā*) with much the same declension as in Burgenland (and some analogies in terms of vowel length). **Vend** (Hungary; Bodnárová 2015: 171) is similar, but also has *āka|dā |jā |lā*; ^E*kova* (f. *koj*), as well as ^E*oko* (optionally indeclinable!), incorrectly traced back to **oko-va* (Bodnárová 2015: 174) (which yielded only *okā*, not *oko*). Another description of **Vend** (Vekerdi 1984: 70) gives a similar system without *oko* but with variation in m.sg. *adā ~ ado*; *odā ~ odo* and some forms in *d-*: pl. *dala* (for **dāla*), m.obl. *dōle*. **Versend** (Hungary; Bodnárová 2009: 59): *ad|ā |(d')a |ana* и *od|ā |(d')a |ana* with a secondary dir.pl. in *-na* (note ^E*kōva*).

2.9. Central Romani: discussion

Proceeding from our initial schema, the following general trends are apparent. The paradigm **ava*, *aja*, *ala* became deficient due to the contraction *ava* > **ā* (except in Plaščuna; maybe optionally in Bourgeois and Kalina), so that we have:

<i>ad(av)a</i>	<i>ad(aj)a</i>	<i>a(da)la</i>	— West Slovak: Burgeois, Kalina
<i>adā</i>	<i>adi</i>	<i>a(da)la</i>	— Uzh
<i>ada</i>	<i>aja</i>	<i>ala</i>	— Bergitka, West Slovak: von Sowa (?) (+ indecl. <i>*ada</i>)
<i>adā</i>	<i>aja</i>	<i>adala</i>	— Prekmurje, Burgenland, Vend (= the Vend group)
<i>adā</i>	<i>a(d')a</i>	[<i>adana</i>]	— Versend

The parallel discrete series (lost in South Central):

<i>kava</i>	<i>*kaja</i>	<i>*kala</i>	— West Slovak: Kalina
[<i>(a)kada</i>]	<i>(a)kaja</i>	<i>(a)kala</i>	— Bergitka
<i>akā</i>	<i>aki</i>	<i>akāla</i>	— Uzh (pl. <i>kāla</i> is now part of <i>kadā</i>)

3. Summary and assessment

Merging dialect data for each group, we obtain:

Proto-North-East:

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–				*jo v j n(e)		
<i>d-</i>		<i>*da</i>	<i>*da va ja la</i>	<i>*do</i>		<i>*do va ja la</i>
<i>_d-</i>		<i>*ada</i>	<i>*ada va ja la</i>	<i>*odo</i>		<i>*odo va ja la</i>
<i>k_d-</i>		<i>*kada</i>	<i>*kada va ja la</i>	<i>*kodo</i>		<i>*kodo va ja la</i>
<i>_k_d-</i>		<i>*akada</i>	<i>*akada va ja la</i>			
<i>k-</i>		<i>*ka</i>		<i>*ko</i>		<i>*ko va ja la</i>
<i>_k-</i>						
<i>k_k-</i>						

Proto-North-West:

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–				*jo v j n		
<i>d-</i>		<i>*da</i>	<i>*da va ja la</i>	<i>*do</i>		<i>*do va ja la</i>
<i>_d-</i>		<i>*ada</i>	<i>*ada va ja la</i>	<i>*<u>ado</u></i>		<i>*odo va ja la</i>
<i>k_d-</i>						
<i>_k_d-</i>						
<i>k-</i>		<i>*ka</i>		<i>*ko</i>		<i>*ko va ja la</i>
<i>_k-</i>		<i>*aka</i>		<i>*<u>ako</u></i>		<i>*oko va ja la</i>
<i>k_k-</i>						

“Proto-Balkan”:

	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–		<i>*a v j l</i>	<i>*a va ja la</i>		<i>*o v j l~n</i>	<i>*o va ja la</i>
<i>d-</i>		<i>*da v j l</i>				
<i>_d-</i>	<i>*ada</i>	<i>*ada v j l</i>	<i>*ada va ja la</i>	<i>*o<u>da</u></i>	?	<i>*odo va ja la</i>
<i>k_d-</i>						
<i>_k_d-</i>						
<i>k-</i>	(<i>*ka</i>)	<i>*ka v j l</i>	<i>*ka va ja la</i>	(<i>*ko</i>)		
<i>_k-</i>	<i>*aka</i>	<i>*aka v j l</i>	<i>*aka va ja la</i>	<i>*o<u>ka</u></i>	<i>*oko v j l</i>	<i>*oko va ja la</i>
<i>k_k-</i>			<i>*kaka va ja la</i>			<i>*koko va ja la</i>

Proto-North-Vlax:						
	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–	?? *av- *ej- **			*vo v j n(e)		
d- _d- k_d- _k_d-	*kad av ej al *kad ava eja ala			*kod ov oj al *kod ova oja ola		
k- _k- k_k-	*k av ej al		*k ava eja ala	*ko v j l		*ko va ja la
		?	*ak ava eja ala	*oko v j l		?
			*kak ava eja ala	*kuk ova oja ola		
Proto-South-Vlax:						
	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–	?? *a- ** **			*(v)o v j n		
d- _d- k_d- _k_d-	*kada	*kad ava eja ala		*koda	*kod ova oja ola	
	*akada			*akoda		
k- _k- k_k-	*ka	*ka va ja la		*ko	*ko va ja la	
		*aka va ja la			*oko va ja la	
Proto-Central:						
	-∅	-PS	-PS-a	-∅	-PS	-PS-a
–			* ? a ja la	*o v j n(e)		* ? o ja la
d- _d- k_d- _k_d-	(?) *ada	*ada va ja la		(?) *oda	*odo va ja la	
		*kada va ja la			*kodo va ja la	
	(?) *akada	*akada va ja la		(?) *okoda		
k- _k- k_k-		*ka va ja la		(?) *oko	*ko va ja la	
		*aka va ja la			*oko va ja la	

The dialect data seem to fit well into the initial schema. The differences among the groups seem trivial (we are probably dealing mostly with losses

and retentions). Interestingly enough, many of the isoglosses seem to fit with the “consensus” classification (i.e. four large groups each with two sub-groups). The expletive *kova*, usually preserved unaltered, is observed in all groups and constitutes a conspicuous Proto-Romani feature.

The primary nature of indeclinable deictics is confirmed by their presence in several groups: all of North-East, some North-West (Sinti, Finnish; residually in WelR), North Central (West Slovak), South Central (Vend, Plaščuna), and South Vlax (Gurbet, Vlaxycko, Servicko). Both expected forms *ada*, *odo* are attested only in NE. In other groups, one of the vowels seems to be secondary:

ada, *odo* (as per initial schema: NE);
ada, *ado* (V₁ = *a*-: Finnish Romani);
ada, *oda* (R = *-a*: most Central);
ado, *odo* (R = *-o*: some Central).

Again, this diversity is reducible to the initial schema. Each deviation from it (*-a-* instead of *-o-* and vice versa) constitutes one step of analogy.

The identity of the personal suffix to the article is confirmed, among other things (such as the oblique forms), by its reconstructed syllabicity in combination with the root **-a-i* > Vlax *-ej* (bearing in mind that our claim regarding Vlax umlaut in the f.sg. demonstrative is not too strong).

The primary nature of non-extended forms is less obvious. However, it is clear that the *-v-* in m.sg. like *akava* has nothing to do with the extension, *contra* Matras (who needs the extension in the protoform to account for *-v-*, quite wrongly). Moreover, at the Common Romani stage, the extension must have been attached rather “loosely” to the personal suffix, since the latter was still syllabic (this is, as per our idea, required for Vlax umlaut), i.e., for instance, **ka-o-a* (later: **kaʷa*²¹ > *kavà*), f. **ka-i-a* (> non-Vlax *kajà*, Vlax **keja*), pl. **ka-(e)l-a* or **ka-l(e)-a* (> *kalà*).

If this is correct, the extension *-à* must have been added quite late, when definite articles were already in use, which points to the Greek-language environment, cf. the model Greek *αὐτός* ‘this’ + *ὁ* (article) + noun. The loss of syllabicity in the article before the extension must be of an even later date. So far, our initial schema has proved its worth, and, if we try to actually reconstruct the Proto-Romani state, the general layout may look like this:

V	C	V	C	R	Article	Extension
<i>*a/o</i> (=R)	<i>*k</i>	<i>*a/o</i> (=R)	<i>*k/d</i>	<i>*a/o</i>	dir. <i>*o, i, (e)/(e)l</i>	<i>*a</i>
					obl. <i>*le(s), la, le(n)</i>	

21 This *-y-* may be seen in Constantinescu *gadaua* (1878: 64), *koua* (79), *akaua* (84), *kodoua* (90).

Examples of use: **d-a o (a) murš* ‘this man’, **o-k-o i (a) čhaji* ‘that very [mentioned] girl’, obl.m. **o-d-o le(s) muršes* ‘that [mentioned] man’, obl.f. **k-a-k-a la čha* ‘this very girl right here’.

The exact behaviour of the combinations of the root *a/o* with the article remains phonetically unclear. For instance, while loss of syllabicity is imaginable in **a-o > *au* (cf. preposition *ka + o > kau, kav, ka* in various Vlax dialects), it is harder to account for in **o-o* (yielding *ov, o* ‘he’ and later Vlax *vo(v)* and Central/North *jo(v)* with later prothesis under stress). The combinations of *a/o* with f.sg. *i* are less problematic (see above), since /i/ becomes a glide more “naturally” than /o/. However, ancient alternations of some sort in the forms of the articles cannot be ruled out: just as f.sg. *i* has a very common variant *e*²², there must have been a m.sg. variant **u* (maybe attested in WelR *ū* alongside *ō*). In any case, initial (pre-)Proto-Romani **o-* and **u-* must have certainly been in some kind of (partially?) complementary distribution.²³

When already univerbated with articles, these combinations may have been perceived as parallel to Greek *αὐτός*, f. *αὐτή* ‘this’ (note the constant part *αὐ-*) used as personal pronouns ‘he’, ‘she’, so that **o-o* (**o-u*), f. **o-i*, pl. **o-l(e)* came to be used in the same way (but only in direct forms). The problem is that, in most dialects, we have **on* (and not **ol*) ‘they’, which can probably be accounted for by the reconstruction of a form **(e)ne/*en(e)* (see § 4).

We now turn to relative chronology. The addition of the extension *-á* (and the subsequent loss of syllabicity of the preceding segment) is clearly posterior to the “2nd Proto-Romani contraction over glide”, e.g. obl.sg. (**čhāpikājā > *čhābijājā > *čhābjā >*) **čhaja > čha* ‘daughter’, cf. the parallel obl.sg. *da* ‘mother’²⁴. On the other hand, the plural in *-á* (non-Vlax *dajá* ‘mothers’, *čhajá* ‘daughters’, Vlax *dejá, čhejá*), looks quite secondary in that it (1) eludes the “2nd Proto-Romani contraction over glide”, (2) displays Vlax umlaut, and (3) eludes the “1st North Vlax contraction” (***dea > **dä*). This may have to do with a (long-lasting?) morphonological constraint prohibiting monosyllabic plurals in *-á* (cf. the omnipresent contracted *phabá* ‘apples’, but only *bajá* ‘sleeves’, and not ***ba*²⁵). Be that as it may, the disturbing fact here is that the plurals ‘mothers’ and ‘daughters’ are the same in both non-Vlax and Vlax (save the umlaut), which seems strange, given their clearly independent

22 E.g., Rakhiv has both: *i* sentence/phrase-initially and *e* elsewhere (Ослон ms.-a).

23 Thus, all verbs from OIA *ō-*, *ava-* have yielded *u-* (and not ***o-*), e.g. *urjel* ‘to put on [clothes]’ < **ōdđh-*, *učharel* ‘to cover’ < **avacchādayati*; from OIA *upari* ‘above’ we have both *opr-* and *upr-*, and, most importantly, the article *o* seems to have yielded *v-* in some words, cf. *vast* ‘hand’ < **u + *(h)ast* < OIA *hāsta-* (see all of these items in ЭСЦЯ).

24 This “2nd Proto-Romani contraction over glide” is quite late, as **daji* ‘mother’ is, most likely, an iranism (see ЭСЦЯ).

25 Yet, on the other hand, Kald. *řa*, an irregular plural of *řyl* f. ‘fart’, but this is etymologically obscure (see ЭСЦЯ).

existence long before the “1st NV contraction”. Hence, we are at a loss in terms of locating the relevant processes chronologically, but the situation may hint at a high degree of uniformity of Proto-/Common Romani at the time when these secondary plural endings were suffixed²⁶, and maybe much later.

However, some unknown phonological factors may have been at play here as well. For instance, one can imagine that the abovementioned secondary pl. *-á* was added so late that it did not trigger “yotation” in feminine nouns and adjectives in *-i* (exactly as the deictic extension *-á* did not at once cause the loss of syllabicity in the article). This may have still been the case in Welsh Romani, cf. WelR *rānī* ‘lady’, pl. *rānīā* (Sampson: 141) vs. Russian Romani *ran’á*, but the problem is that the same syllabicity is found in WelR sg.obl. (acc.) *rānīā*. If we declare the latter secondary (which is supported by WelR acc. *čaiā* ‘daughter’ instead of the expected **čā*, which we see in, e.g. dat. *čakī*, but pl. *čaiā* = sg.obl., secondarily), we may fancy a difference in Proto-Romani between pl.dir. and sg.obl. of the type **rani-a* vs. **ranja* and conjecture that Vlax pl. *dejá* ‘mothers’ comes from **daji-a*, which would account for the umlaut (but, alas, not for the lack of the “1st NV contraction”).

Another (independent) line of reasoning would be that Vlax umlaut is the result of the phonologization of a Proto-Romani allophonic alternation of the type **dāji*, **čhāji*, demonstr. **kā-i* (vs., e.g., *baj* ‘sleeve’) where non-Vlax may have lost the allophone [ā] in favor of the usual [a]. Thus, the suffixing of the extension *-a* to deictics presents some unsolved problems, but it is certainly not too ancient.

4. External comparison

It follows from the above that the basic (“adverbial”: Miklosich, Sampson, see 1.2) deictic elements must have been **a* and **o*, quite distinct from the article (despite the homonymy of the element **o* with the m.sg. article *o*). New Indic languages have (at least superficially) similar genderless deictics, cf. the indeclinable Gujarati *ā* ‘this’ and the much more common *o*-type deictics, e.g. Lahnda, Bengali *o* (Bloch 1934: 198). Their origin is disputed, but their genderlessness and invariability seems to be an innovation with respect to OIA and correspond to the Romani situation.

In Romani, the only element to be deemed direct OIA heritage is the

26 In Proto-Romani (or earlier), the plural of (nearly) all feminine nouns acquired the secondarily ending **-ā* < (?) OIA *-āni* (if so, it was taken over from the neuter; OIA feminine pl. *-āh* would have yielded zero), hence, e.g., *čhib* ‘tongue, language’, pl. *čhibá* (and not ***čhib* < OIA pl. *jihvāh*). Originally, this **-ā* was suffixed to feminine and neuter nouns (e.g., once neuter: *kher* ‘house’, pl. *kherá*), but not to masculine ones (this state is best preserved in North Vlax, cf. Kald. sg. = pl. *řom* ‘Gypsy, husband’, see Ослон 2012; the neuter gender as a morphological category was lost later).

(gendered!) definite article, which, most likely (as it usually happens), goes back to an ancient demonstrative pronoun. Here is what we find in one of the Kalderaš varieties (Ослон 2018: 198), where, as usual, the forms of sg.m.obl. = pl. (for a detailed cross-dialectal description see Boretzky 2000). All these forms, except for sg.f.obl., have variants as part of a morphological distribution (allomorphs: separated by “:”) or occur in free variation (separated by “~”):

	m.sg.	f.sg.	pl.
dir.	<i>o</i> : \emptyset	∂ : <i>j</i> (: <i>i</i>) : \emptyset	<i>le</i> ~ ∂l
obl.	<i>le</i> ~ ∂l	<i>la</i>	<i>le</i> ~ ∂l

For Proto-Romani, we may reconstruct:

	m.sg.	f.sg.	pl.
dir.	* <i>o</i> (~ * <i>u</i> ??)	* <i>i</i> (~ * <i>e</i> ?)	* <i>le</i> (~ * <i>el</i> ?)
obl.	* <i>le</i> (~ * <i>el</i> ?)	* <i>la</i>	* <i>le</i> (~ * <i>el</i> ?)

N. Boretzky leaves the question of the origin of the article unanswered (2000: 56), while Y. Matras wrongly derives it from extended forms (so, e.g. **ova* > *ov* > *o*, **oja* > *oj* > *i*, Matras 2002: 110). We see that, m.sg.dir. and f.sg.dir. (but not pl.dir.!) are suppletive to the rest of the paradigm, exactly as in OIA (cf. Bloch 1934: 200), cf. OIA nom.sg. m. *sá(h)*, f. *sá* vs. gen.sg. m. *tásya*, f. *tásyāh*, dat.sg. m. *tásmai*, f. *tásyai*, nom.pl. m. *té*, f. *táḥ*, etc.²⁷ The direct singular forms of the Romani demonstrative-turned-article are of an obscure origin (obviously, not from the OIA *s*- forms), but cf. Shina (Gilgiti) *o(h)*, f. *e(h)* ‘that’ (Bailey 1924: 23).

The oblique forms must go back to the OIA ones. The change **-t-* > Romani *-l-* requires intervocalic position, which Turner (1928) tried to get around, but in Domari, where **-t-* > *-r-*, the oblique series is exactly as in Romani, save the *e-*: *uhu* ‘this’ (f. *ihī*, pl. *ehe*), obl.sg.m. *eras*, f. *era*, pl. *eran* (Matras 2012: 64, 219). This must go back to the oblique forms of OIA *eṣá(h)* (f. *eṣá*) ‘this’ (deictic particle *e* + *sa(h)*). In Romani the initial vowel is preserved in the variant obl.m.sg./dir.pl. *el*, alternating, at least in Vlax, with *le*, which may have, at some point, been phonologically regular, depending, e.g., on the next consonant (resonant or strident).

The oblique endings are transparent. In the MIA ancestor of Romani (as well as Domari), pronominal declension was restructured on the analogy of the noun: gen.f.sg. *tásyāh* ⇒ **tāyāh* (cf. gen. *ṭṣnāyāh* ‘thirst’, see Bubenik

27 This is Indo-European heritage, cf. also the Greek (Koine) article: nom.sg. m. *ó*, f. *ἡ* vs. gen.sg. m. *τοῦ* (Homer *τοῖο*), f. *τῆς*, dat.sg. m. *τῷ*, f. *τῇ*, but again nom.pl. m. *οἱ* f. *αἱ*, etc.

1996: 94), gen.pl.m. *téṣām*, f. *tāṣām* ⇒ **tānām* (cf. gen.pl. *devānām* ‘gods’), Hence:²⁸

OIA			Domari	Romani
m.gen.	<i>(e)tásya</i>	>	m.obl. <i>eras</i>	<i>(*e)le(s)</i>
f.gen.	<i>(e)tásyāḥ</i> ⇒ <i>*(e)tāyāḥ</i>	>	f.obl. <i>era</i>	<i>(*e)la</i>
pl.gen.	<i>(e)téṣām</i> ⇒ <i>*(e)tānām</i>	>	pl.obl. <i>eran</i>	<i>(*e)le(n)</i> ²⁹

At the same time, Romani pl.dir. *le* (or *el*) may directly continue OIA pl.nom. *eté* ‘these’. The personal pronoun 3.pl. **on* instead of **ol* may be conjectured to contain OIA *ena-*, MIA pl. *ne* (see Bloch 1934: 198).

Domari also has the remote series (*uhu*, f. *ihi*, pl. *ehe*), obl.sg.m. *ōras*, f. *ōra*, pl. *ōran*, exactly parallel to Romani *oles*, *ola*, *olen*, where *o-* may be the same basic deictic element.

As for the “deictic consonants” *k-* and *d-*, the latter is obscure, but the former coincides with the interrogative element³⁰.

In many Romani dialects, there are also direct forms of personal pronouns in *l-*, used only as clitics: m. *lo*, f. *li* (secondarily NV *la*), pl. *le*. Matras views them as old personal pronouns (2002: 101), which may be correct, but they certainly look like a secondary generalization of the oblique stem to eliminate suppletivism (i.e. **etaka-*, f. **etikā*).

5. Conclusions

The general conformity of the reconstructed systems to our “initial schema” allows us to state with some certainty that:

- (1) indeclinable demonstratives (also with prefixes) are original;
- (2) indeclinable demonstratives are turned into inflected ones by the suffixing of “personal suffixes” (which are identical to the article); these combinations must have been free before undergoing univerbation;
- (3) the extension *-a* (of unknown origin; only in direct forms) is a later addition, which implies that forms without it may not be derived from those with it.

28 These analogies may also be required, e.g., for Hindi/Urdu, cf. without analogy obl.sg.m. Hindi/Urdu *is* < OIA *etasyah*, but with analogy pl.obl. *in* < **etānām* (Oberlies 2005: 23).

29 The shortening in gen.sg. *-asya* (? > **-āssa*) > **-āsa* > *-es* and gen.pl. *-ānām* > **-ānā* > *-en* seems to have been regular in (unstressed?) endings, cf. the monosyllabic (and, thus, stressed?) Romani obl. *kas* ‘whom’, *man* ‘me’.

30 Whatever the explanation, this could shed light on the origin of *so* ‘what’ (cf. OIA *sa(h)* ‘this’).

Abbreviations

E	only used as an expletive	NV	North Vlax
(E)	also used as an expletive	NW	North-Western
Kald.	Kalderaš	OIA	Old Indo-Aryan
MIA	Middle Indo-Aryan	PS	personal suffix
NE	North-Eastern	WelR	Welsh Romani

References

- Baghbidi, H.R. 2003. The Zargari language. An endangered European language in Iran. *Romani Studies. Fifth Series.* 13 (2): 123–48.
- Baranja, Samanta. 2013. *Amari čhib. Naš jezik. Slovnica prekmurske romščine.* Ljubljana: Pedagoški inštitut.
- Bailey, George. 1924. *Grammar of the Shina language.* London: Royal Asiatic Society.
- Beníšek, Michael. 2017. *Eastern Uzh varieties of North Central Romani (PhD thesis).* Praha.
- Bloch, Jules. 1934. *L'indo-aryen. Du Veda aux temps modernes.* Paris: Librairie d'Amérique et d'Orient Adrien-Maisonneuve.
- Bodnárová, Zuzana. 2009. *Gramatický náčrt romského dialektu maďarské obce Versend (dipl. práce).* Praha.
- . 2015. *Vend Romani: a Grammatical Description and Sociolinguistic Situation of the so-called Vend dialects of Romani (PhD thesis).* Praha.
- Boretzky, Norbert. 1986. *Zur Sprache der Gurbet von Priština (Jugoslawien).* Giessener Hefte zur Tsiganologie. 3 (1–4): 195–216.
- . 1993. *Bugurdži, deskriptiver und historischer Abriß eines Romani-Dialekts.* Wiesbaden: Harrasowitz.
- . 1994. *Romani: Grammatik des Kalderaš-Dialekts mit Texten und Glossar.* Wiesbaden: Harrasowitz.
- . 1996. *Arli: Materialien zu einem südbalkanischen Romani-Dialekt.* Grazer Linguistische Studien. 46: 1–31.
- . 1998. *Erlí. Eine Bestandsaufnahme nach den Texten von Gilliat-Smith.* *Studii Romani.* 5-6: 122–160.
- . 2000. *The definite article in Romani dialects.* In: Elšík, Viktor, and Matras, Yaron (eds.), *Grammatical Relations in Romani. The Noun Phrase.* Amsterdam: Benjamins. 31–63.
- . 2003. *Die Vlax-Dialekte des Romani. Strukturen – Sprachgeschichte – Verwandtschaftsverhältnisse – Dialektkarten.* Wiesbaden: Harrasowitz.
- Boretzky, Norbert, Cech, Petra, and Iгла Birgit. 2008. *Die Südbalkanischen Dialekte (SB1) und ihre innere Gliederung. Analyse und Karten.* Graz: Institut für Sprachwissenschaft.
- Bubenik, Vit. 1996. *The Structure and Development of Middle Indo-Aryan Dialects.* Delhi: Motilal Bandarsidass.
- Bourgeois, Heinrich. 1911. *Kurze praktische Grammatik der mitteleuropäischzigeunerischen Sprache.* Striegau: Huß.
- Calvet, George, Delvoye, Françoise, and Labalette, Michèle. 1970. *Abrégé grammatical du dialecte manuš. Etudes Tsiganes.* 1970(1): 69–79.
- Calvet, George. 2009. *Dictionnaire tsigane-français, dialecte kalderash.* Paris: L'Asiathèque.
- Cech, Petra. 2006. *Dolenjska Romani: The dialect of the Dolenjski Roma in Novo Mes-*

- to and Bela Krajina, Slovenia. München: Lincom Europa.
- Cech, Petra, and Heinschink, Mozes F. 1999. *Sepečides-Romani: Grammatik, Texte und Glossar eines türkischen Romani-Dialekts*. Wiebaden: Harrassowitz.
- Ćirković, Svetlana, and Mirić, Mirjana. 2017. *Romsko-srpski rečnik knjaževačkog gurbetskog govora*. Knjaževac: Narodna biblioteka „Njegoš“.
- Constantinescu, Barbu. 1878. *Probe de limba și literatura țiganilor din România*. București: Tipografi'a societății academice române.
- Constantinescu, Barbu. 2016. *Cântece țigănești. Romané ghilea*. ed. critică, introducere, traducere în limba engleză, tabel cronologic, index și bibliografie de Julieta Rotaru; pref.: Viorel Cosma. București: Editura Muzeul Literaturii Române.
- Elšík, Viktor. 2000. Dialect variation in Romani personal pronouns. In: Viktor Elšík and Yaron Matras (eds.). *Grammatical Relations in Romani. The Noun Phrase: 65–94*.
- Elšík, Viktor, Hübschmannová, Milena, and Šebková, Hana. 1999. The Southern Central (*ahi*-imperfect) Romani dialects of Slovakia and northern Hungary. In: Halwachs, Dieter, and Menz, Florian (Hrsg.). *Das österreichische Romani im europäischen Kontext*. Klagenfurt: Drava. 277–390.
- Finck F. N. 1903. *Lehrbuch des Dialekts der Deutschen Zigeuner*. Marburg: N.G. Elwert.
- Gjerdman Olof, and Ljungberg Erik. 1963. *The language of the Swedish coppersmith gipsy Johan Dimitri Taikon; grammar, texts, vocabulary and English word-index*. Uppsala: A.-B. Lundequistska Bokhandeln; København: Ejnar Munksgaard.
- Granqvist, Kimmo. 2009. Demonstratives in Romani: A Minimalist account. *Biolinguistica Fennica Working Papers 1 (2009): 105–16*.
- Granqvist, Kimmo. 2011. *Suomen romanin äänne- ja muotorakenne*. Helsinki: Suomen Itämainen Seura ja Kotimaisten kielten tutkimuskeskus.
- Halwachs Dieter. W., and Wogg Michael. 2002. *Grammatik des Burgenland-Romani (Roman)* (unter Mitarbeit von Gerd Ambrosch, Katharina Deman, Ursula Glaeser). Oberwart: Verein Roma.
- Holzinger, Daniel. 1993. *Das Romanes: Grammatik und Diskursanalyse der Sprache der Sinte*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Hübschmannová, Milena, Šebková, Hana, and Žigová, Anna. 1991. *Romsko-český a česko-romský slovník*. Praha: Státní pedagogické nakladatelství.
- Igla, Birgit. 1996. *Das Romani von Aija Varvara. Deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.
- Ješina, Josef. 1882. *Romáňi čib čili jazyk cikánský*. Praha: Fr. A. Urbánek.
- Kalina, Antoine. 1882. *La langue des Tziganes slovaques*. Posen: J. K. Župański.
- Kozhanov, Kirill. forthc.-1. *Russian Romani*.
- Kozhanov, Kirill. forthc.-2. *Grammar of the Plašuna dialect*.
- Leggio, Daniele Viktor. 2011. The dialect of the Mitrovica Roma. *Romani Studies. Fifth Series. 21 (1): 57–114*.
- Lípa, Jiří. 1963. *Příručka cikánštiny*. Praha: Státní pedagogické nakladatelství.
- Mānušs, Leksa, Neilands, Jānis, and Rudevičs, Kārlis. 1997. *Čigānu–latviešu–angļu etimoloģiskā vārdnīca un latviešu–čigānu vārdnīca*. Rīga: Zvaigzne ABC.
- Matras, Yaron. 1994. *Untersuchungen zu Grammatik und Diskurs des Romanes. Dialekt der Kelderasa/Lovara*. Wiesbaden: Harrassowitz.
- . 1998. Romani. Deixis and deictic oppositions in discourse: evidence from Romani. *Journal of Pragmatics. 29-4: 393–428*.
- . 1999. The Speech of Polska Roma. *Journal of the Gypsy Lore Society. Fifth Series. 9 (1): 1–28*.
- . 2000. The structural and functional composition of Romani demonstratives. In:

- Elšík, Viktor, and Matras, Yaron (eds.). Grammatical Relations in Romani. The Noun Phrase. Amsterdam: Benjamins. 95–112.
- . 2002. Romani. A linguistic introduction. Cambridge: University Press.
- . 2003. Romacilikanes – The Romani dialect of Parakalamos. *Romani Studies. Fifth Series.* 14 (2): 59–109.
- . 2012. *A Grammar of Domari.* Berlin/Boston: De Gruyter Mouton.
- Miklosich, Franz Xaver Ritter von. 1872–80. Über die Mundarten und Wanderungen der Zigeuner Europas, 12 Tle. Wien.
- Minkov, Michael. 1997. A concise grammar of West Bulgarian Romani. *Journal of the Gypsy Lore Society. Fifth Series.* 7 (2): 55–95.
- Oberlies, Thomas. 2005. *A historical grammar of Hindi.* Graz: Leykam.
- Oslon, Mikhail. 2017. Jek pušimos la istorijaka fonetikako la kêldêraricko šibaki: le vokalongê žutengi *e : ê aj i : î* distribucija. In: Kozhanov, Kirill, Oslon, Mikhail, and Halwachs, Dieter W. (red.). *Das amen godi pala Lev Čerenkov. Romani historija, čhib taj kultura.* Graz: Grazer Linguistische Monographien: 325–38.
- Paspati, Alexandre. 1870. *Études sur les Tchinghianés; ou, Bohémiens de l'Empire otoman.* Constantinople: Impr. A. Koroméla.
- Pobožniak, Tadeusz. 1964. *Grammar of the Lovari dialect.* Kraków: Państwowe Wydawnictwo Naukowe.
- Romlex. Lexical Database: romani.uni-graz.at/romlex.
- Rozwadowski Jan M. 1936. *Wörterbuch des Zigeunerdialekts von Zakopane.* Kraków: Polska Akademia Umiejętności.
- RMS = Romani Morpho-Syntax Database: romani.humanities.manchester.ac.uk/rms.
- Sampson, John. 1926. *The Dialect of the Gypsies of Wales Being the Older Form of British Romani preserved in the Speech of the Clan of Abram Wood.* Oxford: University Press.
- Schmidt, Ruth Laila, and Kohistani, Razwal. 2008. *A grammar of the Shina language (in collaboration with Mohammad Manzar Zarin).* Wiesbaden: Harrassowitz.
- Sechidou, Irene. 2011. *Balkan Romani: The dialect of Ajjios Athanasios/Greece.* München: Lincom Europe.
- Sowa, Rudolf von. 1887. *Die Mundart Der Slovakischen Zigeuner.* Göttingen: Vandenhoeck und Ruprecht.
- Tensor, Anton. 2005. *Lithuanian Romani.* München: Lincom Europa.
- . 2008. *Northeastern group of Romani dialects (PhD thesis).* Manchester.
- Turner, Ralph L. 1928 *Romani les and Sanskrit tásya.* *Bulletin of the School of Oriental Studies, University of London.* 5 (1): 43–51.
- Vekerdi, József. 1984. The Vend Gypsy dialect in Hungary. *Acta linguistica hungarica.* 34/1-2: 65–86
- Wagner, Peter. 2012. *A Grammar of North West Lovari Romani (PhD thesis).* Praha.
- Windfuhr, Gernot L. 1970. European Gypsies in Iran: A First Report. *Anthropological Linguistics* 12. 271–92.
- Αλεξίου, Ιωάννης Γ. Λεξικό της Ρομανί γλώσσας (ελληνορομανό). Online edition: repository.edulll.gr/edulll/retrieve/2350/692.pdf.
- Деметер, Р.С., and Деметер П.С. 1990. *Цыганско-русский и русско-цыганский словарь (кэлдэрарский диалект), под ред. Л.Н. Черенкова.* Москва: Русский язык.
- Кантя, Г. В. 1970. *Фолклорос романо. Кишинёв: Карта молдовеняскэ.*
- Кожанов, К.А. forthcoming-3. *Грамматический очерк диалекта цыган-кишинёвцев.*
- Кожанов, К.А. forthcoming-4. *Грамматический очерк языка цыган-сэрвов.*
- Ослон, М.В. 2012. *Отражение древнеиндийского среднего рода в цыганском.*

- Вопросы языкового родства. № 8. Москва, 2012. 93–100.
- Ослон, Михаил. 2014. О неким особинама ромског нагласног система. IWoBA VIII: реферати VIII Међународног скупа о балтословенској акцентологији (Нови Сад 2012). Славистички зборник, нова серија, књига I: 293–315.
- Ослон, М.В. 2018. Язык котляров-молдовая. Грамматика кэлдэрарского диалекта цыганского языка в русскоязычном окружении. Москва: ЯСК.
romanes.org/pub/Oslon/Ослон М.В. Язык котляров-молдовая.pdf
- Ослон, М.В. forthc.: Короткий опис волоського діалекту циган з м. Рахів.
romanes.org/pub/Oslon/Ослон М.В. Короткий опис волоського діалекту циган з м. Рахів.pdf
- Панченко. Я.А. 2013. Изучение влашского диалекта ромского языка: поиски и находки. Киев: Научно-исследовательский центр ориенталистики им. О. Прицака НАУКМА.
- Смирнова-Сеславинская, М.В., Цветков, Г.Н. 2009. Грамматика цыганского языка (кэлдэраря и ловаря на основе языка ловаря). Межкультурная русско-цыганская коммуникация. Москва: Институт национальных проблем образования.
- Торопов, В. Г. 2003. Словарь языка крымских цыган. Москва.
- ЭСЦЯ: Ослон М.В., Кожанов К.А. Этимологический словарь цыганского языка:
romanes.org/pub/ЭСЦЯ.pdf