Variation in the nominal morphology of Northern Vlax Romani

Márton A. Baló
Research Institute for Linguistics, Eötvös Loránd Research Network, Budapest, Hungary

---

Abstract

The present paper discusses two particular instances of variation in the nominal morphology of Northern Vlax Romani varieties as spoken in Hungary: the masculine oblique base and the feminine plural oblique base. The discussion is conducted in an analogical framework, relying only on surface forms and their relationships, using the notion of schemas (Booij 2010), and taking it one step further. When there is a "weak point" in the grammar of a language, variation may emerge and pattern-seeking may begin; the pattern-seeking processes can be interpreted and explained with reference to possible analogical connections among surface forms.

Keywords: variation, analogy, patterns, paradigms, Construction Morphology, Vlax Romani

1. Introduction

This paper discusses two particular instances of variation in the nominal morphology of Northern Vlax Romani varieties as spoken in Hungary, focussing on variation in the strict sense, that is, phenomena involving vacillating stems, and tries to explain them by locating and identifying the possible analogical sources of the variation.

The first instance is the masculine paradigm, which displays two distinct patterns for the formation of the oblique stem: *-es-* in the singular and *-en-* in the plural on the one hand, and *-os-* in the singular and *-on-* in the plural on the other. The appearance of these two patterns is fairly predictable under certain conditions but the choice becomes uncertain when more factors are involved. The second instance is the feminine paradigm, which shows an even higher degree of variation, and the uncertainty about the choice of the oblique plural pattern, *-an-* or *-en-*, is the result of two competing pressures: the uniformity of the oblique patterns within each gender and the uniformity of the oblique plural across both genders.

The discussion is conducted in an analogical, exemplar-based framework, relying only on surface forms and their relationships. Exemplar representations are built up of rich and detailed information about the tokens encountered through linguistic experience, including variation and features that may otherwise be predictable (Bybee 2013: 52). The memories are organised in a cognitive map of categories based on similarity (Pierrehumbert 2001: 140), and categorisation takes place on all linguistic levels and in all domains, which also implies that "linguistic exemplars come in a variety of sizes, ranging from a single segment, such as a vowel, to whole paragraphs" (Bybee 2013: 53), and anything in between, including morphological units such as the Romani oblique markers.

This process of categorisation leads to the emergence of constructions; therefore, the discussion will also rely on Construction Morphology, drawing on the notion of schemas containing phonological, syntactic and semantic information, as introduced by Booij (2010), and taking this notion one step further by suggesting that schemas possess a circular rather than purely linear structure and by combining schemas to show their interconnectedness.

My hypothesis is that when there is a weak point in the grammar of a language, variation may emerge. A weak point may arise if different morphological patterns are employed by different stems for the same semantic function within a strictly delimited paradigm. If these patterns enter into competition, the weak point becomes a locus for the emergence of morphological variation. This competition may cause a stem to employ varying means of expressing the same function or to altogether change its means of expressing that function.

2. Background and data

In this section, I provide a brief introduction to Romani, its speakers and the classification of its dialects, followed by an overview of existing corpora and the data collected during the research.

2.1 Background

Romani is a New Indo-Aryan language spoken primarily in Europe, its closest relatives among other Indo-Aryan languages being Rajasthani and Gujarati. Although realistic estimates of the number of speakers are not easy to make, Bakker (2001) puts their number at approximately 4.6 million in Europe at the beginning of the third millennium. The total number of the Romani people in Europe, including both speakers and non-speakers, has been estimated at anywhere between 4 and 12 million. Due to more recent migrations, Romani has also been spoken in the Americas, where the numbers are even harder to determine, but conservative estimates (cf. Matras 2005a) suggest that there are upwards of 500 000 speakers, and there are probably more, as there are about 800 000 Romani people living in Brazil alone (cf. Gaspar 2012) and approximately one million in the United States (Hancock 2013). Romani monolingualism is virtually non-existent; probably all speakers of Romani are at least bilingual, save for some children who live in communities that are fairly isolated from the local non-Romani population. Their bilingualism is mostly unidirectional, which means that they acquire the language of the surrounding majority population but Romani is only seldom acquired by the members of the majority population; therefore, "Romani dialects exhibit a range of contact influences from different languages" (Matras & Adamou 2020: 329-330). However, oblique markers, which form the central focus of the present paper, are never borrowed (Elšík & Matras 2006: 234); therefore, bilingualism is unlikely to have played a role in the facts at issue here.

Vlax Romani is classified as one of the four main dialect groups of Romani based on a shared set of language-specific innovations. Miklosich (1872-80) mostly relied on contact phenomena, and more specifically on the European loan elements in the lexicon for his classification of Romani dialects, while another influential proposal put forward by Gilliat-Smith (1915) was the distinction between Vlax and Non-Vlax dialects based on diagnostic features (Matras 2002: 218-9). These classification schemes have given way to a more state-of-the-art theory in the current literature on Romani linguistics, which builds upon shared features and internal innovations (Matras 2005b) in order to establish the separate dialects and dialect branches. Based on certain diagnostic features, Vlax is further divided into a Southern and a Northern branch, with the former spoken mostly in the Balkans, the latter in Romania, Hungary, Moldova and Serbia. Some confusion is caused by the fact that the Northern Vlax group has often been referred to simply as Vlax Romani in the literature on Romani published in Hungary (e.g. Erdős 1959; Vekerdi 2000).

2.2 Existing corpora and the newly collected data

Naturalistic and trustworthy corpora of any one variety of Romani, did not exist until very recently, and when looking into instances of synchronic variation, new and naturalistic data are of utmost importance. If we look at the international landscape, the situation is certainly improving, although the small corpora of Thrace Romani-Turkish-Greek and Finnish Romani-Finnish have been collected with the aim of research into language contact (Adamou 2016), and the corpus of Russian Romani does not yet include newly collected spoken data (Kozhanov 2016). The linguistic atlas of Central Romani <http://ulug.ff.cuni.cz/atlas/> documents cross-dialectal variability in Central Romani, while the two largest and most ambitious Romani databases to date, ROMLEX <http://romani.uni-graz.at/romlex/> and the Romani Morpho-Syntax Database <http://romani.humanities.manchester.ac.uk/rms/> are dialectologically overarching. Although they do not focus on one specific dialect, they are worth consulting when doing research into any dialect.

ROMLEX is a lexical database that contains a large amount of data representing the variation in the lexicon of all Romani dialects; the RMS Database is a comparative description of Romani dialects with the principal focus on morphosyntax, but some lexical and phonological features are also included.

With that in mind, the decision was made to collect new, naturalistic, up-to-date Northern Vlax Romani data. Based on a questionnaire, fieldwork was conducted in several locations in Hungary. The questionnaire was specifically designed to focus on the masculine oblique base and the feminine oblique base. It contains 204 Hungarian sentences in random order, which were read out to the consultants, who were asked to translate them into Romani. A recording was made of the interview, which was then transcribed. The transcriptions used for the purposes of this paper focussed on the lexical items that are analysed below. More thorough and detailed transcription of all of the material in the ELAN[1] format is under way.

For the masculine oblique base, the questionnaire focussed on two kinds of lexical items. The first group consisted of words where variation was suspected; these included words in the concise dictionary of Romani dialects in Hungary (Vekerdi 2000), such as *čókano* ‘hammer’, *dúhano* ‘tobacco’, *búso* ‘bus’, *čalā́do* ‘family’, *kirā́ji* ‘king’, *sókro* ‘father-in-law’, *fő́ro* ‘town’, *trájo* ‘life’, and *pohā́ri* ‘glass’. The second group contained lexical items that are supposedly not part of the lexicon as such, so consultants had to provide a translation on the spot, e.g. *láptopo* ‘laptop’, *móbilo* ‘mobile phone’, *pokrő́co* ‘blanket’, and *teléfoni/teléfono* ‘telephone’. Here, our expectation that these words would inflect according to the masculine paradigm proved to be right. The sentences were formed so as to contain inflected forms of the target nouns, since these would reveal their oblique patterns. Besides the targeted items, we also acquired data about selected items of the Indo-Aryan vocabulary. For the feminine plural oblique base, no differentiation was made within the vocabulary, as variation seems to affect the whole of the feminine paradigm. The targeted items included every-day words like *pīrí* ‘saucepan’, *mesají* ‘table’, *katt* ‘pair of scissors’, *patrí* ‘leaf’, and *bórotva* ‘razor’; and words for poultry and insects, such as *cincā́ri* ‘mosquito’, *ʧīrí* ‘ant’, *māčí* ‘fly’, *khajnʲí* ‘hen’, and *papín* ‘goose’. The sentences were generally designed to reflect possible every-day usage and to sound as natural as possible. Some examples are shown in Table 1, with target words in italics. The individual target words are listed in the relevant sections of the paper, where more information is provided concerning the number and frequency of tokens.

Table 1: Some examples of the sentences from the questionnaire, used for testing the nominal oblique patterns

Original Hungarian sentence	English translation	Romani translation
Elmentem a *családdal* a városba.	I went to town *with the family*.	Gēlém tar e *čalādósa* ándo fő́ro.
Leesett a *kalapácsnak* a feje.	The head *of the hammer* came off.	Tḗle pēlás e *čokanósko* šēró.
Egész nap a *mobiljával* játszik.	He plays *with his mobile* phone all day.	Sōró djes e *telefonésa* khelél pe.
Eljöttem az *asztaloktól*.	I came away *from the tables*.	Avilém tar e *mesajéndar*.
Tele van *szúnyogokkal* a szoba.	The room is full *of mosquitoes*.	Pherdó-j *cincārénca* i sóba.

For the purposes of the current paper, 30 sets of sentences were analysed, coming from interviews with native speakers of one or another variety of Northern Vlax Romani between the ages of 28 and 82, living in the counties of Baranya, Fejér, Hajdú-Bihar, Heves, Komárom-Esztergom, Pest, Somogy, Tolna, Szabolcs-Szatmár-Bereg and Veszprém. The quantity of data is not large (amounting to approximately 30,000 tokens at present), and not every target word was elicited from

[1] ELAN is an annotation tool for audio and video recordings, freely available at https://archive.mpi.nl/tla/elan. It was developed by the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands; for more detail, see Sloetjes & Wittenburg 2008.

every native speaker. Even so, small corpora have been used effectively in the course of conducting valuable research. Adamou (2016) mentions a study of Welsh-English bilinguals (Deuchar 2006) based on a five-hour corpus of which only forty minutes were analysed (Adamou 2016: 15). Larger corpora exist for more widely spoken languages, as is the case for French-English bilinguals in Canada (Poplack 1993). In the present study, I rely exclusively on these data, but nevertheless occasionally refer to another sufficiently reliable but slightly outdated source, Vekerdi (1985). The Northern Vlax Romani varieties reflected in these data include Lovari, Mašari and Drizari. Although sometimes considered as separate linguistic groups within Northern Vlax Romani (Erdős 1959; Tálos 2001), they are not treated as separate dialects by Boretzky (2003) in his comprehensive study of the Vlax dialects of Romani. Based on their similarity seen so far, they will be considered as constituting a single variety by linguistic criteria, but used by different, self-designated groups.

3. Variation, weak points and schemas

In this section, I explain what should be meant by variation in the present paper and then define the notion of a weak point in relation to this particular form of variation.

3.1 Variation

The term macrovariation has been used to characterize differences across languages, while microvariation denotes variation within a language (Dufter et al. 2009b). The latter term, however, frequently refers to inter-dialectal variation, as seen for example in Barbiers et al. (2007), who also explicitly note that they are concerned "with an idealised language system and not with dialect internal variation" (Barbiers & Cornips 2007: 3). This statement reflects the fact that the papers in the volume in which it appears deal with syntactic variation, and, besides a typological approach, adopt a generative stance. Other research into microvariation tries to account more "for the range and (limits) of inter- and intra-speaker variation in a principled way while at the same time testing existing formal theories against these microvariational data and thus contributing to the theory of language variation", but it still investigates closely related language variants, thus adopting a dialectologically oriented approach, "applying the formal theoretical concepts of generative grammar" (Brandner 2012: 113). While such an approach is certainly an improvement compared to a previous state in which variation was altogether left to the field of sociolinguistics, it still takes on a geographically, socially or stylistically oriented perspective (cf. e.g. Neef 2009). At the same time, a fresh take on variation is provided by the same volume, claiming that "even totally unconditioned ('free') variation exists", but they also add that "variation is not something we expect a grammar to produce" (Dufter et al. 2009a: 9).

Since the publication of Dufter et al. (2009b), the study of free variation in morphology has attracted renewed interest, as e.g. in Rainer et al., eds., (2010b), where it is defined as several patterns competing for the same kind of base, "a one-to-many relationship between a meaning and the exponents serving to express that meaning" (Rainer et al. 2010a: 4). They add that "competition is rampant in natural languages, especially in word formation" (Rainer et al. 2010a: 4), and, although much less, in inflection as well, as Paster (2010) demonstrates through the example of Maay, a Cushitic language spoken in Somalia, where "all consonant-final nouns have three possible plural forms: one form with the suffix *-o*, one form with the suffix *-yal*, and a third form with both suffixes" (Paster 2010: 177). The last option, in which multiple realisations of a single feature occur within a word, is that of multiple exponence (Harris 2017). This phenomenon in the nominal morphology of Maay is not unlike the case in Romani, although, as we will see below, the optionality of the variation in Romani is probably not as straightforward as in Maay. Dammel & Schallert (2019b) attempt to fill the lacuna in research into variation in inflection, presenting diverse instances of inflectional variation "in different guises and domains: within a single cell of a paradigm and across paradigms, within one spatially defined variety and across varieties,

synchronically and diachronically, within and across speakers" (Dammel & Schallert 2019a: 1). They also remark that "morphology has not been in the centre of dialectological (or variationist linguistic) interest" (Dammel & Schallert 2019a: 3). While this may be true for Germanic varieties explored in their volume, Romani has been studied extensively as far as dialectological variation is concerned (two of the most recent and most comprehensive sources being the RMS Database mentioned above and Boretzky & Igla 2004; but specific phenomena have also been investigated by, among others, Elšík 2000a; van den Heuvel & Urech 2014; Leggio & Matras 2017).

A considerable merit of Dammel & Schallert (2019b) is, however, that they draw attention to the naturalness of morphological variation, its significance for linguistic theory, the importance of empirical evidence for morphological theory and the possible need for different degrees of granularity when considering variation. The present study is in line with these considerations insofar as it takes variation as an essential feature of any language variant and it attempts to link empirical data to morphological theory. Romani data are especially valuable in this respect, given that "one of the most fascinating aspects of the Romani linguistic landscape is its great diversity" (van den Heuvel & Urech 2014: 43). The variation discussed below is of an intra-dialectal nature and is treated as such in the paper. However, as it also shows signs of intra-speaker variation, a finer and more granular analysis of the data and a more detailed look at other factors such as the age and geographical location of consultants may, in the future, provide further insight into its nature.

3.2 Weak points and schemas

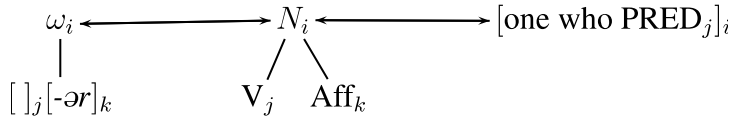
A weak point is understood as the locus of the emergence of variation. It can fundamentally be described as a point "in the paradigm where more than one conflicting analogical requirement applies with approximately equal strength" (Rebrus & Törkenczy 2011: 139).[2] The significance of a weak point lies in the fact that although choice can lead to variation, it does not necessarily do so. Paradigms and individual members of paradigms can be quite rigid, like the closed noun class of lowering stems and lowering suffixes in Hungarian. When followed by a suffix, lowering stems and lowering suffixes require a low linking vowel instead of the regular mid vowel. In this case, even if the phonological form of a stem is very similar to that of a non-lowering stem, no variation emerges: *gáz* 'gas' is a regular, non-lowering stem, the plural form of which is *gázok*, while *ház* 'house', a lowering stem has the plural form *házak*, and no alternative forms (**gázak* or **házok*) ever occur.[3]

In order to expound on the more precise nature of a weak point, I use the idea that regularities on a particular level of linguistic description can be expressed in terms of schemas as defined in the word-based framework of Construction Morphology (Booij 2010, following the notion of schemata, as described by Rumelhart 1980). Morphological schemas are closely related to constructions, which denote a pairing of form and meaning (Goldberg 1995; Jackendoff 2008), and different from rules as they do not involve a derivational phase; rather, they are declarative in nature (Jackendoff & Audring 2016), containing phonological, morphosyntactic and semantic information and representing a system of connections between these components. Schemas are ultimately lexical items, having "the same format as words, differing only in that part of their structure consists of variables" (Jackendoff & Audring 2016: 471). See, for example, the schema for English deverbal *-er* in Figure 1, where the symbol ↔ stands for correspondence (Booij 2010: 8).

[2] Although the present paper mainly deals with formal connections, Rebrus & Törkenczy (2011) add that a functional relationship can also serve as an analogical connection.

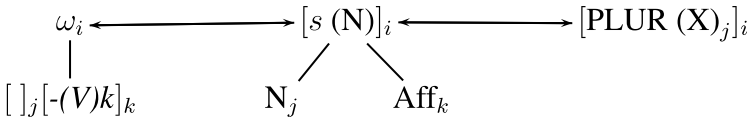
[3] Some well-known examples of cases where choice does lead to variation and ultimate change in the long term can be found in the English past tense, which will be touched upon again further below in this section. Old English *bacan* 'bake', for instance, was a strong verb but started to weaken in late Middle English and eventually became weak, as opposed to *wacan* 'wake', which remained strong.

Figure 1. Schema for deverbal *-er* in English



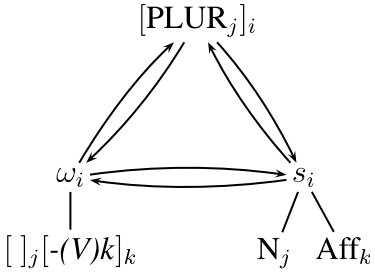
The three kinds of linguistic information included here are the phonological form, the semantic information and the morphosyntactic information. The morphosyntactic information in the original form of the schema is encoded as N, meaning that the word containing the deverbal *-er* suffix is a noun. However, there may often be a need for morphosyntactic properties to be specified (Booij 2010: 7). The schema for the Hungarian plural suffix *-k* may be represented as in Figure 2, containing variables linked by the co-index *j*. The morphosyntactic component is labelled *s*, indicating that a schema can represent constructions of any level of morphological or syntactic complexity and can be further expanded to include more detailed information about its morphosyntactic properties, which, in this case, is the fact that the resulting plural item is a noun.

Figure 2. Schema for the Hungarian plural suffix *-k*



Instead of this linear representation, which is based on the idea of Booij (2010), I suggest a circular representation of the schema, as sketched in Figure 3, where all three pieces of information are connected to the other two through correspondences. This arrangement is justified by the relationship that exists between the semantic and the phonological information, which is not unlike what Jackendoff (2012) suggests, for example, when he claims that a word like *cow* is stored in memory, and "it involves a pronunciation linked with a meaning and the syntactic feature *Noun*" (Jackendoff 2012: 176). One reason why it is important to postulate interrelations among all three components is, as it is pointed out by Jackendoff (2012), that there are words which lack one of the components, like *ouch,* which has phonology and meaning but lacks syntactic features. Another reason for postulating a direct relationship between the phonological and the semantic components is that this particular relationship is a significant one in the argumentation below: as I show, the variation seen in the nominal morphology of Northern Vlax Romani takes place along the phonology/semantics correspondence while leaving the syntactic component intact. This provides sufficient evidence for the solution proposed here, a schema showing an interrelated matrix of the three components.

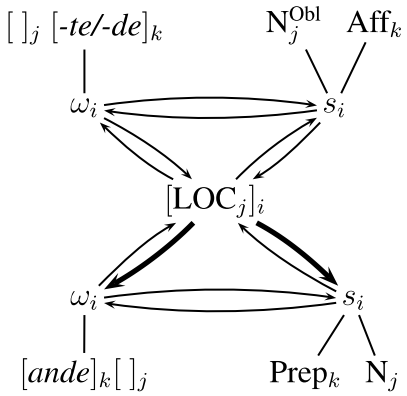
Figure 3. Improved schema for the Hungarian plural suffix *-k*



A schema like this becomes weaker when there is a disturbance in any of the correspondences. For example, if a new phonological form ω'$_i$ started to appear in the same syntactic position and with the same meaning as the deverbal *-er* or the plural *-k*, then this would weaken the schema, which means that it would cease to be the exclusive plural marker, losing its all-powerfulness. This in turn may trigger variation and the schema would become a weak point. It is also possible that more than one correspondence becomes unstable, like the locative case in Northern Vlax Romani, where the semantic component may pair up with a different phonological form and a different syntactic position, resulting in variation. Thus, a weak point in morphology is a schema where at least one of the correspondences is not biunique.[4]

The representation in Figure 4 combines two schemas for the locative case in Northern Vlax Romani. The upper section of the combined schema represents the agglutinative case marking (that of *e kheréste* 'in the house'): it specifies phonological form; the morphosyntactic information that the case affix is attached to the oblique base of the noun; and the corresponding locative semantics. The lower section of the combined schema presents an alternative way of expressing the locative, by means of a preposition, as in *andó kher* 'in the house'.[5]

Figure 4. Schema for the locative case in Northern Vlax Romani



We now come to the significance of the replacement of the single, bidirectional arrow with two distinct, unidirectional arrows: the directions can be highlighted separately by thickening only one arrow within one correspondence. This is meant to signify that the direction in question is stronger than any of the other directions in any of the correspondences in the combined schema. In this particular case, it means that the prepositional form in the expression of the locative case is more common than the agglutinative form. However, the presence of both forms suggests that the locative function does not exclusively correspond to either the form represented by agglutinative case marking or the form represented by the preposition.

As another example, let us take the English past tense. There is a strong relationship between the semantic function "past tense" and its most regular expression, the addition of the suffix *-ed*. If all English verbs inflected that way, there would only be a single schema for the past tense.

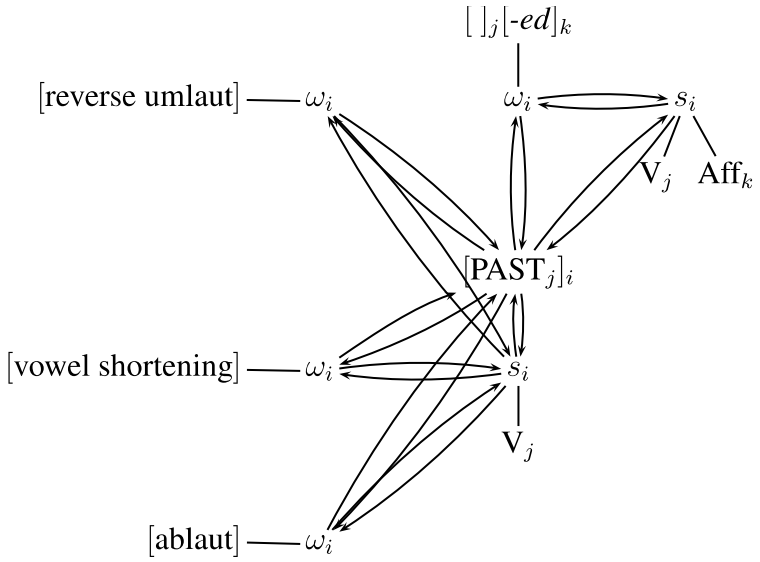
However, this is not the case. There are several alternative, so-called "irregular" verbs of lower or higher frequency, making up smaller or bigger groups (*sing-sang*, *cut-cut*, *keep-kept*, and so on). The existence of these groups of verbs also means that the correspondence between the past

[4] The one-to-one correspondence between semantics and syntax (more specifically between the participant roles associated with a verb and the argument roles associated with a construction) is well-known in construction grammar (Goldberg 1995: 51). The lack of biuniqueness discussed here is different from those cases when "a verb's meaning and the meaning of the argument structure construction it occurs in are not necessarily in a biunique correspondence" because this simply refers to the fact that "the same verb meaning can appear in more than one argument structure construction" (Goldberg 2013: 446).

[5] The form is in fact made up of the preposition *andé* and the definite article *o*. The article immediately precedes the noun, as seen in the agglutinative form of the locative, and inflects for case, gender and number; hence the form *e*.

tense function and the marker *-ed* is not inevitable, and neither is the correspondence between the past tense function and the morphosyntactic property of affixation for the past tense. Several other morphosyntactic patterns and phonological forms are used in the formation of the English past tense, for example ablaut (*sing-sang*), vowel shortening (*keep-kept*) or reverse umlaut (*think-thought*).

Figure 5. Schema for the English past tense



With so many schemas coalescing around the same semantic component, the correspondences become ambiguous and represent a weak point, where variation may emerge, although it does not necessarily do so. This may depend on other factors, such as the frequency of the different ways of forming regular and irregular past tense forms or the degree to which individual word forms are entrenched in the speaker's memory (see Blything et al. 2018 for an account in favour of the single route model as discussed for example by Bybee & Moder 1983 in children's acquisition of English past tense forms). Although there is an increasing amount of empirical data to support the relationship between frequency and entrenchment, "the distinctly usage-based claim that high token frequency results in the entrenchment of fully compositional form-meaning pairings is extremely controversial" (Blumenthal-Dramé 2012: 23). If variation does emerge, though, then we have every reason to think that there are patterns which are competing for the same function, or patterns which have some other kind of phonological or morphosyntactic influence on the forms that begin to vary.

4. An overview of the weak points under discussion in Northern Vlax Romani

In this section, I briefly introduce the two weak points in the nominal inflection of Northern Vlax Romani where variation occurs and where the surface forms (surface similarities and differences; in general, cf. e.g. Kálmán et al. 2012) and analogical effects might play a role in producing and maintaining what seems to be variation of an intra-dialectal nature as described in Section 3.1.[6]

1. The first weak point to be examined is the masculine oblique base. One oblique marker for masculine nouns is *-es-* in the singular and *-en-* in the plural, so the oblique bases of a word like *šēró* 'head' are *šērés-* and *šērén-*, respectively. However, this marker is not the only one used in the masculine paradigm. The schema is weakened by the existence of another set of phonological

[6] In most cases, this also proved to be intra-speaker variation. Although some speakers showed consistency in their use of the two patterns, not applying the *-es-/-en-* marker to borrowed nouns, these cases were few and far between, and no regularities in age or geographical location were observed.

forms, *-os-* in the singular and *-on-* in the plural, so, for example, the oblique forms of the word *híŕo* 'a piece of news' are *hīrós-* and *hīrón-*, respectively.
2. The second weak point can be found in the feminine plural oblique base. The oblique marker in the singular is invariably *-a-*: *šej* 'girl' ~ *šejá-*, *žuv* 'louse' ~ *žuvá-*. However, there are two available patterns in the plural. One of the possible feminine plural oblique markers is *-an-*: the plural oblique base of *šej* 'girl' is *šeján-*; but there is another feminine plural oblique marker, *-en-*, as in *žuvén-*, the plural oblique base of *žuv* 'louse'.

The present paper does not deal with the diachronic processes that have led to the existence of the different patterns. However, it must be noted that the two masculine patterns *-es-/-en-* and *-os-/-on-* have been claimed to descend from an earlier distinction between two declension classes—respectively, the so-called thematic or oikoclitic class for inherited nouns and the athematic or xenoclitic class for borrowed nouns. For discussion of the diachronic processes, see primarily Bakker 1997; Boretzky 1989; Elšík 2000b; see also Elšík & Matras 2006: 72 for a list of reconstructed Early Romani declension classes, including several subclasses. The feminine plural oblique marker *-an-* is modelled on the oblique singular in *-a-* (Matras 2002: 83), replacing the original form in *-en-*. There is overwhelming diachronic evidence for the strict morphological split between vocabulary inherited from Indo-Aryan (including early loans from Persian, Armenian and Greek) and lexical items borrowed at a later time from Greek and other contact languages (Romanian, Serbian, Hungarian etc.); Elšík & Matras 2006 refer to this split as chronological compartmentalisation by. However, interaction between the two layers has already been noted (Elšík 2000b: 23; Elšík & Matras 2006: 327-332), and the fact that the current state of affairs manifests variation indicates that a strict dichotomy no longer holds on a synchronic level. One reason for this may be that there have not been two specific and unique morphological systems used for the two separate layers; their inflection, strictly taken, is not different.[7]

For the sake of the present study, I refer to each alternating set of oblique markers as a pattern. Associating each pattern with a declension class can run the risk of circularity (especially in the case of the feminine); by the same token, the evidence for or against assuming that a given masculine form is borrowed isn't always clear-cut. Nevertheless, if we consider them in terms of canonical inflection ("a unique mapping from form to function, and from function to form"; Corbett 2009: 1) and overabundance (the phenomenon wherein "there are two or more forms available to realise the same cell in an inflectional paradigm"; Thornton 2012: 183), then the nouns with canonical paradigms, whose oblique forms do not vary, may be said to belong to one or the other of two different declension classes, while varying nouns may be considered as having overabundant paradigms in which an oblique cell is occupied by competing forms from distinct declension classes.

5. The masculine oblique base

In this section, we will look at the first weak point, the masculine oblique base, in more detail. Following the description of the phenomenon in question, we will analyse two possible reasons for the weakness and the ensuing variation, and discuss to what extent there can be interaction between the possible reasons and the variation.

5.1 Description of the phenomenon

[7] Extralinguistic factors might also interfere in the oblique form, although the extent to which a native speaker considers a certain word to be borrowed or not is very difficult to measure. As Elšík & Matras (2006) also note, "this compartmentalisation is not based on specialists' knowledge about the origin of the lexemes"; rather, "it is directly encoded in the structure of the language and, at least in principle, accessible to its speakers" (Elšík & Matras 2006: 324). One might expect the nominal classes to remain absolutely rigid, with no variation at all; however, this is not the case: as the inherited or borrowed nature of a word is becoming obscure without direct access to the donor language, there appears to be more hesitation.

In this section, we will introduce the variation in the masculine oblique base and we will also see that this variation is closely interlinked with the masculine nouns with a stem-final /o/. In Northern Vlax Romani, there are two sets of suffixes for the oblique base.

(1) a. *šērró* 'head' → *šērés-/šērén-* OBL
b. *hír̀ro* 'a piece of news' → *hīrós-/hīrón-* OBL

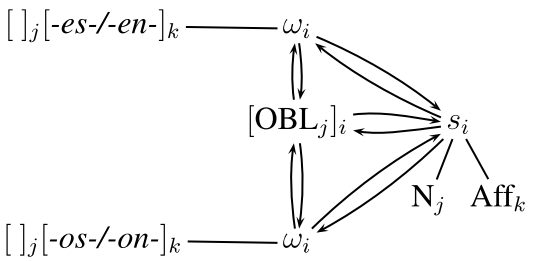
The first set comprises *-es-* for the singular and *-en-* for the plural, but there are masculine nouns which, without any apparent phonological or morphophonological reason, exhibit a different oblique pattern: *-os-* in the singular and *-on-* in the plural. While the vowel /e/ is very common across the whole inflection, the vowel /o/ only appears in the masculine oblique markings *-os-* and *-on-*. Table 2 illustrates the two competing patterns throughout the whole paradigm.

Table 2: The two masculine paradigms in Northern Vlax Romani

	bakró 'sheep'		*sókro* 'father-in-law'	
	Singular	Plural	Singular	Plural
Nominative	bakró	bakré	sókro	sokrurá
Accusative	bakrés	bakrén	sokrós	sokrón
Dative	bakréske	bakrénge	sokróske	sokrónge
Locative	bakréste	bakrénde	sokróste	sokrónde
Ablative	bakréstar	bakréndar	sokróstar	sokróndar
Instrumental	bakrésa	bakrénca	sokrósa	sokrónca
Genitive	bakrésk-	bakréng-	sokrósk-	sokróng-
Vocative	bákra	bakrále	sókra	sokrále

Masculine nouns can be divided into three groups based on the distribution of the oblique patterns: in the first group, only the oblique pattern in *-es-/-en-* is used; in the second, only the oblique pattern in *-os-/-on-* is used; and in the third, the two possible patterns vary. The masculine oblique base might therefore be represented as the schema in Figure 6,where N is a masculine noun. This schema contains both patterns, one where the phonological form of the oblique marker is *-es-/-en-* and another where the phonological form is *-os-/-on-*. Although the oblique case is ultimately a morphological category, it can be seen as corresponding to a syntactic and a semantic component. The syntactic component S_i accounts for an oblique noun's syntactic position and structure, while [OBL] represents the semantic content of the oblique case. In Conceptual Semantics (cf. e.g. Jackendoff 1991), plurality, for example, is encoded as a function in the semantic structure. The oblique can also be considered a function whose domain is the set of nouns; the value of this function for a given noun is the aggregate of the oblique forms of that noun, which can thus serve as the domain of other functions representing the nominal cases.

Figure 6. Schema for the masculine oblique base



In this combination of two separate schemas, one containing the phonological form $\omega_i[\ ]_j[es/en]_k$ and the other one containing the phonological form $\omega_i[\ ]_j[os/on]_k$, the same semantic content corresponds to two different phonological forms. The correspondence between the phonological form $\omega_i[\ ]_j[es/en]_k$ and the semantic content OBL_j in the first schema is weakened by the presence of the second schema, where the same semantic content corresponds to a different phonological form, $\omega_i[\ ]_j[os/on]_k$. The presence of the first schema likewise weakens the phonology/semantics correspondence in the second schema.

As illustration, consider the masculine nouns found in the newly collected data; these are listed in Tables 3–5. Only items which have at least one attested oblique form are included here. Table 3 contains 29 masculine nouns with oblique forms exhibiting the pattern *-es-/-en-*; Table 4 contains 24 masculine nouns whose oblique forms exhibit the *-os-/-on-* pattern. In addition to these, seven lexical items whose oblique forms vary between the two patterns are listed in Table 5. Here, two word forms that realize the same oblique case/number combination, one conforming to the *-es-/-en-* pattern and the other to the *-os-/-on-* pattern, can be considered cellmates (two forms realising the same cell, as defined by Thornton 2012), and the locus of the variation is the paradigm cell to which they correspond. For the sake of greater transparency, the words in each table are grouped together according to their number of syllables. Within each such group, words are listed according to the final sound of the nominative singular form, which can be /i/, /o/ or a consonant.

Table 3: Masculine nouns with the oblique pattern *-es-/-en-* in the questionnaire dataset

Nominative singular	Attested oblique forms
	one syllable
berš ‘year’	*beršésko*
del ‘god’	*devléske, dēvléske*
gad ‘shirt, clothes’	*gādénca, gadéske, gādéske, gādénge, gādéngo*
gav ‘village’	*gavéske*
grast ‘horse’	*grastéske, grastén*
kašt ‘tree’	*kaštéske, kaštésa, kašténge, kašténca*
kher ‘house’	*kheréske, kherésko*
kraj ‘king’	*krajéske, krajénge*
murš ‘man’	*muršéske*
nāj ‘finger’	*nājénca*
rom ‘Romani man’	*roméske, roménca, romén, romés*
than ‘place’	*thanéste, thanés*
vast ‘hand’	*vastésa*
	two syllables
abáv ‘wedding’	*abavéske*
bijáv ‘wedding’	*bijavéske*
gurúv ‘bull’	*guruvén*
kotór ‘cloth’	*kotorésa*
manúš ‘man’	*manušés, manušén, manušéste, manušésko, manušéstar, manušénca, manušénge*
bāló ‘pig’	*bālén*
gāžó ‘non-Romani man’	*gāžéske, gāžéstar, gāžén*
kurkó ‘week’	*kurkéstar*
šāvó ‘boy’	*šāvéske, šāvés, šāvén, šāvénge, šāvénca*
	three syllables
gẃezeši ‘train’	*gēzešésa*
kirā́ji ‘king’	*kirājéske, kirājénge, kirājén*
koldúši ‘beggar’	*koldušéstar, koldušés, koldušén, koldušénca*
kopā́či ‘tree trunk’	*kopāčéske*
pohā́ri ‘glass’	*pohārénca*

Table 4: Masculine nouns with the oblique pattern *-os-/-on-* in the questionnaire dataset

Nominative singular	Attested oblique forms
	two syllables
ā́tko ‘curse’	*ātkónca*
búso ‘bus’	*busósa*
čā́so ‘hour, watch’	*čāsóngo*
fṓro ‘town’	*fōróske*
gíndo ‘problem’	*gindóstar, gindónca*
hī́ró ‘a piece of news’	*hīróstar*
nā́so ‘child’s father-in-law’	*nāsósko*
nī́po ‘relatives’	*nīpósa, nīpós*
pújo ‘chicken’	*pujón*
ríto ‘field’	*ritóske*
sókro ‘father-in-law’	*sokróske, sokrónge*
trájo ‘life’	*trajósko*
	three syllables
āláto ‘animal’	*ālatón, ālatós*
bārṓvo ‘baron’	*bārōvóske*
čalā́do ‘family’	*čalādós, čalādósa, čalādón*
falató ‘a little bit of food’	*falatóske, falatón*
xāmásko ‘food’	*xāmaskós*
jōsā́go ‘livestock’	*jōsāgós*
laptópo ‘laptop’	*laptopósa*
sómsēdo, sómsī́do ‘neighbour’	*somsēdósko, somsēdóski, somsīdós, somsēdós, somsēdóske*
vonáto ‘train’	*vonatósa*
	four or more syllables
ternimā́ta ‘the young ones’	*ternimātós, ternimātóske, ternimātónca, ternimātónge*
šegīččḗgo ‘help’	*šegīččēgóske, šegīččēgós*
sā́mītōgḗpo ‘computer’	*sāmītōgēpósa*

Table 5: Masculine nouns exhibiting variation between the *-es-/-en-* and *-os-/-on-* patterns in the questionnaire dataset

Nominative singular	Attested oblique forms
	three syllables
bašadó ‘telephone, mobile’	*bašadésa, bašadósa*
čókano ‘hammer’	*čokanésko, čokanósko*
dúhano ‘tobacco’	*duhanés, duhanéski, duhanós, duhanóski*
mobiló ‘mobile phone’	*mobilésa, mobilósa*
pokrőco ‘blanket’	*pokrōcésa, pokrōcósa*
	four syllables
kirčimári ‘bartender’	*kirčimārésa, kirčimārósa, kirčimāréstar, kirčimāróstar, kirčimārénca*
telefóni, telefóno ‘telephone’	*telefonésa, telefonósa*

As for the stems whose oblique case forms vary, the variation, which is located within the paradigm cells corresponding to those cases, is slight in some instances, with one or the other pattern more dominant, but there are instances, like *dúhano*, where we find that the occurrences of the two different oblique patterns are almost equal in number.[8] The overall proportion of the frequency of the stems with the *-es-/-en-* pattern, those with the *-os-/-on-* pattern, and those whose patterns vary can be seen in Table 6.

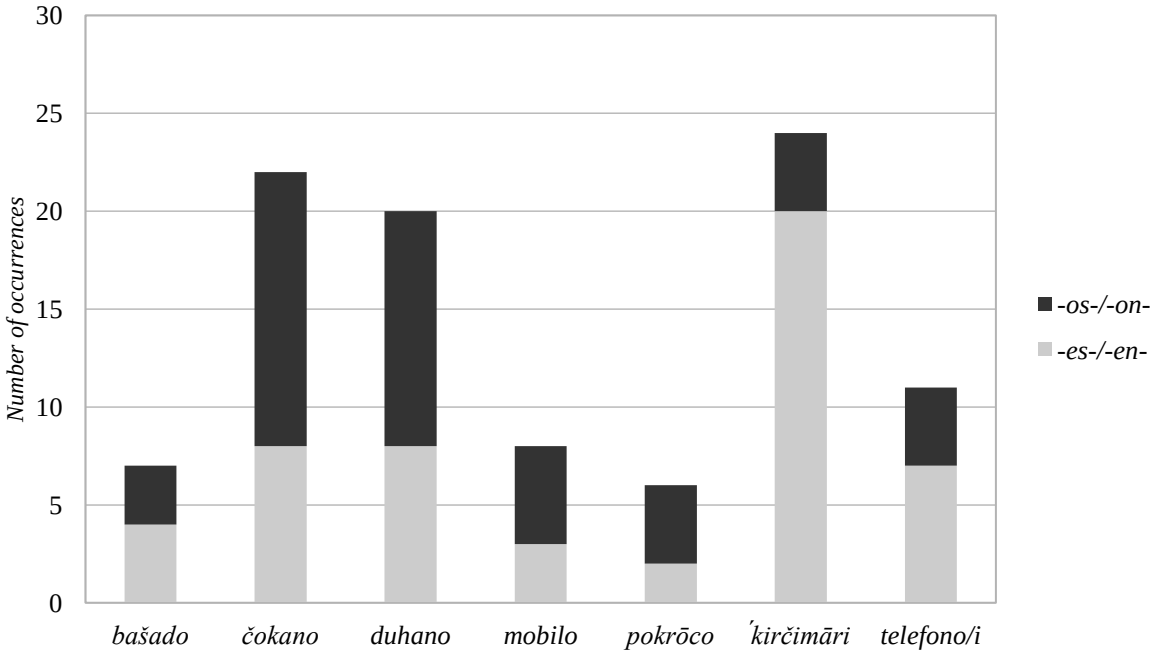
Table 6: Number and proportion of the frequency of the stems with the *-es-/-en-* pattern, the *-os-/-on-* pattern, and the alternating pattern in the questionnaire dataset

oblique pattern	number	percentage
-es-/-en-	29	48%
alternating	7	12%
-os-/-on-	24	40%

The varying stems and the total number of occurrences of both variants in the data are repeated in Figure 7. The number of varying stems in the masculine oblique is seven out of a total number of 60 nouns.

[8] More evidence for the variation comes from Cech et al. (1999), who provide a further example: the word *kókalo* ‘bone’ has two possible oblique forms, *kokalós-* and *kokalés-*.

Figure 7. The total number of occurrences of the varying masculine stems in the questionnaire dataset



5.2 Possible causes and explanations

Table 5 and Figure 7 show that variation seems to appear more often among words where the final vowel of the nominative singular form is /o/: see for example *čokáno* ‘hammer’, *dúhano* ‘tobacco’, *mobílo* ‘mobile phone’. As can be seen in Table 3, masculine nouns ending in /i/ take the oblique in *-es-/-en-*. The word *telefóno* apparently has an alternative nominative form, *telefóni*, and there are some other masculine nouns ending in /i/ which show variation, such as *kirčimā́ri* ‘bartender’. The fact that the oblique form of the word *telefóni/telefóno* ‘telephone’ appears both as *telefonés-* and as *telefonós-* might as well be the result of the different nominative forms. Instances of two different nominative forms have been attested elsewhere, for example *mūšoró* and *mūšorí* ‘programme’. With regard to the variation in the word *kirčimā́ri* ‘bartender’, it must be noted that there are ambiguous cases as the nominative of the same word takes other forms in the dataset as well, such as *kočmāróši* and *kirčimārúši*, but there is not enough information available to draw a conclusion from them. While in the Kalderaš dialect, Boretzky (1994) documents oblique forms with *-es-/-en-* only for nouns with a stem-final /i/, for example *limóri* ‘grave’ ~ *limorés-/limorén-*, Cech & Heinschink (1999) only quote masculine nouns with a stem-final /i/ where the oblique suffixes are *-os-/-on-*, for instance *juhā́si* ‘shepherd’, obl. *juhāsós-*; *doktóri* ‘doctor’, obl. *doktorós-*; and so on. The newly collected Northern Vlax Romani data from Hungary corroborate the findings of Boretzky (1994). We now examine two possible explanations for the variation.

1. Position of stress. At first glance, it seems there is at least some sort of correlation between the variation of the oblique forms and the fact that Northern Vlax Romani, for historical and contact linguistic reasons, lacks a straightforward stress pattern. Stress itself appears to vary, especially in words with three syllables. While the stress pattern of disyllabic words (word-initial or word-final) seems to determine the form of the oblique base unambiguously, the varying stress pattern of trisyllabic words frequently entails the unpredictability of oblique forms.
2. Number of syllables. This is related to the position of stress to some degree, as oblique forms begin to vary when the number of syllables reaches or exceeds three. The variation is especially

ostensible on trisyllabic words with a stem-final /o/, while disyllabic words, as mentioned above, never vary.

5.2.1 *Variation in the position of stress*

In this section, we examine the relationship between the variation in the position of stress in the nominative and the appearance of one or the other oblique pattern. The choice of words in Table 2 (*sókro* 'father-in-law' and *bakró* 'sheep') suggests that the oblique form is the direct consequence of the nominative stress pattern. We will see that there is certainly a correlation between the two aspects, but it is far from being so straightforward: besides the stem-final vowel, stress is only one of several factors that might play a role in the formation of the oblique form.

Generally, and especially for disyllabic words, where the stress falls on the last syllable of the nominative singular form, there is no variation: the oblique pattern is *-es-/-en-*. On the other hand, where the stress falls on the first (penultimate) syllable of the nominative singular, the oblique pattern is *-os-/-on-*. No matter what the oblique ending is and where the stress falls in the nominative singular form, the stress in the oblique forms always falls on the oblique ending, so *trájo* 'life' gives the oblique stem *trajós-*. In inflected forms, this results in penultimate stress: dative *trajóske*, locative *trajóste*, ablative *trajóstar* and instrumental *trajósa*.

For words with three syllables, on the other hand, stress may vary widely. While the oblique suffix attracts stress, and so the forms inflected for case always have penultimate stress, stress in the nominative forms can fall on any of the three syllables.

As we can see in Tables 3-5, the position of stress cannot unambiguously predict the oblique form. While it is true that words with stem-final stress take the oblique pattern *-es-/-en-* without exception, the oblique form of words where the stress falls on the penultimate or ante-penultimate position is not so obvious. The words *padlővo* 'floor' and *rablővo* 'robber, highwayman', for example, have the oblique forms *padlōvés-* and *rablōvós-*, respectively (cf. Vekerdi 1985), in spite of the fact that both bear stress on the penultimate syllable. The newly collected data show that the oblique form of certain stems varies, for example *mobílo* 'mobile phone', obl. *mobilés-/mobilós-* or *dúhano* 'tobacco', obl. *duhanés-/duhanós-*. The choice of pattern may be further complicated by the fact that the stress of the nominative form may even vary within one stem, for example *kóčiši/kočíši* 'coachman' (Vekerdi 1985). In sum, where stress varies in the nominative, the vowel of the oblique suffix also varies.

5.2.2 *The number of syllables*

There might be a correlation between the number of syllables in a noun and the degree of variation that it shows with respect to oblique forms. In this section, we examine this correlation, eventually coming to the conclusion that the more syllables a stem has, the more likely it is to exhibit variation in its oblique pattern.

Monosyllabic nouns always end in a consonant and invariably take the *-es-/-en-* oblique pattern, as in *drom* 'road', obl. *dromés-*. This pattern is valid for disyllabic nouns ending in a consonant, e.g. *rašáj* 'priest', obl. *rašajés-*. The *-os-/-on-* pattern appears when two factors present themselves simultaneously: disyllabicity and a stem-final vowel. The stem-final vowel, which eventually gets deleted, introduces a certain amount of disturbance in the system because it conflicts with the initial vowel of the oblique suffix. Among disyllabic stems with a stem-final /o/, however, there is no variation in the strict sense: every lexical item which has two syllables and a stem-final /o/ exhibits either one or the other pattern, and the position of the stress (final or penultimate) appears to be a reliable clue in this case, as seen in Tables 3-5.

When the number of syllables increases to three, variation emerges on the level of lexical items. This means that the longer a word is, the more uncertain it is which oblique pattern it will employ. As mentioned above, there is only slight variation for words longer than two syllables which end in a different vowel, such as /i/: the frontness of the stem-final vowel dominantly

predicts (or triggers) a front vowel in the oblique form. The back vowel /o/ of nouns with three syllables, however, is not able to predict the oblique form unambiguously, just as the nominative form of disyllabic nouns ending in /o/ cannot predict the oblique form.

Although there must be some among the trisyllabic masculine nouns with a stem-final /o/ that take *-es-/-en-* as their oblique (as attested in Vekerdi 1985, for example), our newly collected data do not present such items. However, they present nine items with the oblique pattern *-os-/-on-* and five items whose oblique pattern varies. This is somewhat in line with the varying stress pattern of trisyllabic nouns: the increase in the number of syllables enhances the chance of variation, too. While the oblique pattern of disyllabic nouns never varies (it is either *-es-/-en-* or *-os-/-on-*), it begins to vary when the number of syllables exceeds two. This is further corroborated by the two items with four syllables, *kirčimáŕi* 'bartender' and *telefóni/telefóno* 'telephone', especially if we consider the proportion of varying nouns with four syllables to non-varying ones: there are only three items that show the oblique *-os-/-on-* pattern and there are none that show the oblique *-es-/-en-* pattern.

In connection with the higher number of *-os-/-on-* oblique forms we should also note that when variation begins, that is, at the level of trisyllabic nouns, the stem-final /o/ might tip the scales in favour of the oblique *-os-/-on-* pattern (whereas the word *kirčimáŕi* 'bartender', with a stem-final /i/, seems to prefer the *-es-/-en-* pattern).

5.3 Interim summary

In this section, we had a look at the first weak point in the morphology of Northern Vlax Romani, the masculine oblique base, in more detail. Following the description of the phenomenon in question, we went over two possible reasons for the weakness and the ensuing variation (which only emerges if the stem-final vowel is /o/) and we found the following.

1. The position of stress. We saw that the stress pattern of disyllabic words (word-initial or word-final) corresponds to the choice of the oblique marker: word-initial stress corresponds to *-os-/-on-*, word-final stress corresponds to *-es-/-en-*. Stress begins to vary in trisyllabic words, and the same lexical item can occur with different stress patterns. That is exactly where the oblique markers begin to vary: if there is a varying stress pattern, the choice of the oblique pattern becomes unpredictable.
2. The number of syllables. We found that while the oblique forms of disyllabic nouns do not vary, the oblique forms of trisyllabic nouns with a stem-final /o/ do. Based on this, it seems that the number of syllables influences the choice of oblique forms: the higher the number of syllables is, the higher the possibility of variation is.

6. The feminine oblique plural base

In this section, we look at the second weak point, the feminine oblique plural base, in more detail. Following the description of the phenomenon, we examine two possible aspects that might influence the choice of the plural oblique ending for feminine nouns.

6.1 Description of the phenomenon

The feminine oblique singular base has one single pattern: *-a-*, so the oblique form of *šej* 'girl' is *šejá-*. However, there are two possible alternatives for the oblique plural base: one is *-an-*, so the oblique plural base of a word like *khajnʲí* 'hen' is *khajnʲán-*, but there is another one, *-en-*: for instance, the oblique form of *rắca* 'duck' is *rācén-*. They occur simultaneously as the feminine oblique plural base in several cells of the feminine paradigm. Diachronically, the *-an-* pattern arises through the analogical influence of the oblique singular in *-a-* (Matras 2002: 83; Elšík 2000b: 22; Boretzky 1994: 33); this change reinforces the formal differentiation of the feminine and masculine paradigms. For example, the oblique plural base of *krangá* 'branch' is *krangán-* in Hungarian

Lovari (Hutterer & Mészáros 1967: 49), from an original oblique plural in *-en-*, and this analogical process is said to have taken place most often in the Vlax dialects. However, the plural oblique of *krangá* ‘branch’ exclusively appears as *krangén-* in the newly collected data, which suggests that the old form has been retrieved after an intermediary stage of renewal.

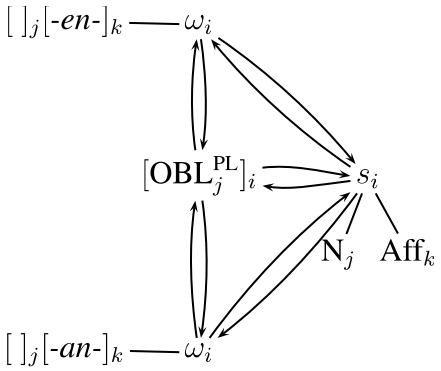
The two feminine oblique plural bases are again treated as competing patterns. Similarly to the masculine oblique forms, alternative forms in *-an-* and *-en-* realizing the same oblique case/number combination are cellmates (cf. Thornton 2012), and the locus of the variation is again the paradigm cell to which they correspond. Table 7 exemplifies the two different feminine paradigms. Note that the oblique singular forms of feminine nouns are completely unaffected by variation: the singular oblique marker is invariably *-a-*.

Table 7: The two different patterns in the feminine in Northern Vlax Romani

	rắca ‘duck’		*māčí* ‘fly’	
	Singular	Plural	Singular	Plural
Nominative	rắca	rācí	māčí	māčá
Accusative	rācá	rācén	māčá	māčán
Dative	rācáke	rācénge	māčáke	māčánge
Locative	rācáte	rācénde	māčáte	māčánde
Ablative	rācátar	rācéndar	māčátar	māčándar
Instrumental	rācása	rācénca	māčása	māčánca
Genitive	rācák-	rācéng-	māčák-	māčáng-
Vocative	rắca	rācále	mắča	māčále

The two different patterns can be represented by the combination of two schemas in Figure 9, where N is a feminine noun. The correspondence between the phonological form $\omega_i[\ ]_j[an]_k$ and the semantic content OBL $PLUR_j$ is weakened by the presence of the other schema, where the same semantic content corresponds to a different phonological form, $\omega_i[\ ]_j[en]_k$. Conversely, we can also say that the correspondence between the phonological form $\omega_i[\ ]_j[en]_k$ and the semantic content OBL $PLUR_j$ is weakened by the presence of the other schema, where the same semantic content corresponds to a different phonological form, $\omega_i[\ ]_j[an]_k$.

Figure 8. The combination of two schemas for the feminine oblique plural



The feminine nouns from the newly collected data can be seen in Table 8. The items are grouped together according to their oblique plural forms; items with no attested plural oblique form are excluded. Out of a total of twenty items, four items take the oblique plural marker *-an-*, seven take the oblique plural marker is *-en-*, and there are nine stems where the oblique forms vary. A striking fact here is that **t**he number of stems where there is variation is much highe**r** than would have been expected based on earlier sources, such as Vekerdi (1985).

Table 8: Feminine nouns and their oblique forms in the questionnaire dataset

Nominative singular	Attested oblique forms	
nouns with the oblique plural in *-an-*		
xajíng ‘well’	*xajingánge, xajingángo*	
khajnʲí ‘hen’	*khajnʲán*	
pīrí ‘saucepan’	*pīránge*	
māčí ‘fly’	*māčánca*	
nouns with the oblique plural in *-en-*		
angrustí ‘ring’	*angrusténdar*	
	armajá ‘curse’	*armajénca*
	cincā́ri ‘mosquito’	*cincārénca*
	kangrí, krangí ‘branch’	*kangrénca, krangénca*
	kúrva ‘whore’	*kurvéngo*
	mesají ‘table’	*mesajéndar*
rā́ca ‘duck’	*rācén*	
nouns with variation		
katt ‘pair of scissors’	*kattʲánca, kattʲénca*	
	māj ‘meadow’	*māján, mājánge, mājénge*
	papín ‘goose’	*papinʲán, papinʲén*
	patrí ‘leaf’	*patrénca, patránca*
	šūrí ‘knife’	*šūránca, šūrénca*
	tʲīrí ‘ant’	*tʲīránca, tʲīrénca*
	bāj ‘sleeve’	*bājánca, bājénca*
bār ‘garden’	*bāránge, bārán, bārénge*	
bórotva ‘razor’	*borotvénca, borotvánca*	

The overall proportion of the frequency of the stems belonging to the two feminine oblique plural patterns and the stems where the oblique forms vary is given in Table 9. These proportions show that the feminine class of nouns is even more affected by variation than the masculine class, with a higher percentage of all the attested stems showing variation.

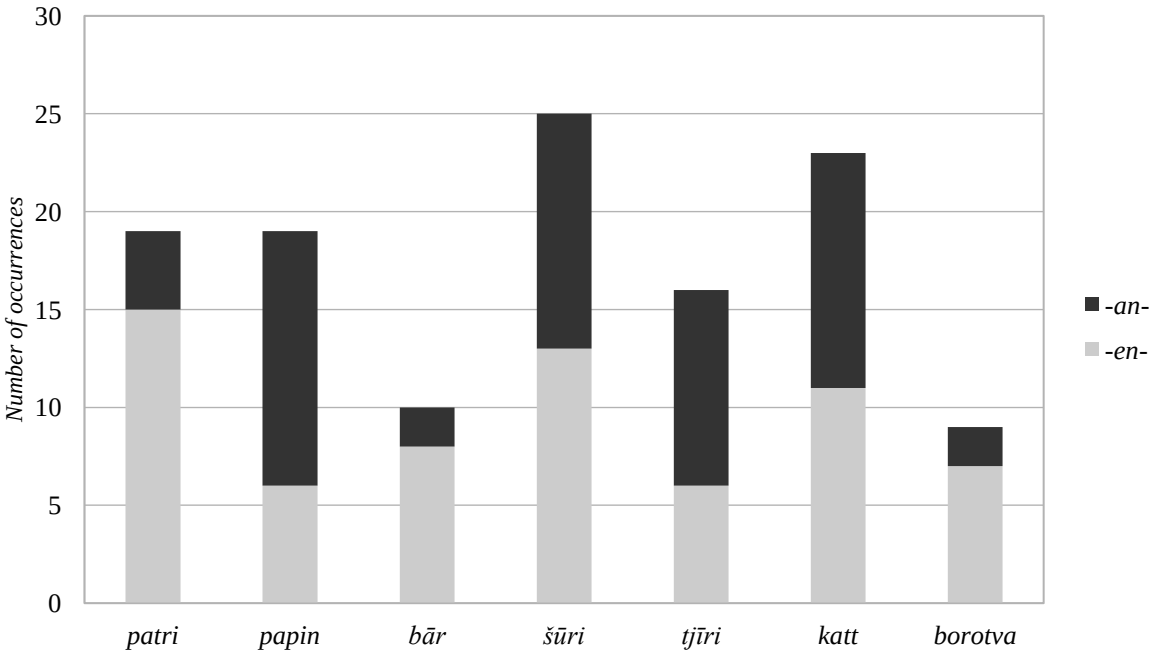
Table 9: Number and proportion of the stems exhibiting the two feminine oblique plural patterns and the varying stems in the questionnaire dataset

oblique pattern	number	percentage
-en-	7	35%
alternating	9	45%
-an-	4	20%

The varying stems and the total number of occurrences of both variants in the data are repeated in Figure 9, except for two items, where the variation is very slight and needs further evidence: only

one instance containing the suffix *-an-* for *bāj* ‘sleeve’ and only one instance containing the suffix *-en-* for *māj* ‘meadow’ were found.

Figure 9. The total number of occurrences of the varying feminine stems in the questionnaire dataset



Cech & Heinschink (1999) try to explain the two different patterns by reference to the difference between inherited and borrowed words: inherited words take *-an-* and borrowed words take *-en-*. This, however, is highly unlikely as it is completely inconsistent with the data (with the exception of *papín* and *bórotva*, all the varying stems are inherited) and even with the way the inherited-borrowed dichotomy in the masculine is traditionally analysed; this explanation should therefore be dismissed.

While the vowels /u/ and /i/ appear less often in suffixes in general, and even then they are more typically used in derivation, the vowels /e/ and /a/ appear quite frequently in the inflection of Romani, for example as the vowel component of the nominal oblique markers, both in the feminine and masculine paradigms, or of the personal concord markers on verbs. This general frequency of /a/ and /e/ and their relative frequency in the Romani verbal and nominal suffixes might be determining factors behind the presence and competition of the two patterns, although this is contradicted by the fact that the proportion of the two different forms varies among the different stems.

As Table 10 shows, the personal concord markers for consonantal verbs (with the inclusion of the /e/, which is analysed as epenthetic by Baló 2008) exclusively contain these two vowels.

Table 10: Verbal personal concord markers in Northern Vlax Romani

	1st sing.	2nd sing.	3rd sing.	1st plural	2nd plural	3rd plural
present	*-av*	*-es*	*-el*	*-as*	*-en*	*-en*
past	*-em*	*-an*	*-as*	*-am*	*-an*	*-e*

If we consider the fact that the first and second person plural forms are less frequent generally (see a summary of cross-linguistics surveys on the frequency of person and number features in child and adult natural speech in Austin 2012: 262-264), we see that the proportion of personal concord markers containing /e/ and /a/ is 5:3, which corresponds to the tendencies we find for the distribution of the two vowels in the feminine oblique plural marker. Even if both the verbal and the oblique markers reflect a more general distribution or proportion of the vowels within the language, it is important to see that the distribution does not only present itself as different nominal classes formed with one or the other vowel, but also as stem-level variation, where one single stem can form the oblique with both markers.

The vowels /e/ and /a/ are both very common in the nominal oblique markers, which, including both the masculine and the feminine paradigms, can be *-es-*, *-en*, *-a-*, *-an-*, while, as mentioned before, the vowel /o/ only appears in the additional variant forms *-os-* and *-on-* in the oblique masculine and nowhere else in the inflection. The variation between *-en-* and *-an-* in the feminine oblique plural is also much more salient than in the masculine, with variation seen in nine stems out of 20 compared to the variation in the masculine oblique between *-es-/-en-* and *-os-/-on*, where the proportion is a mere seven out of 60.

We can compare the proportions seven out of 60 and nine out of 20 by means of a chi-square test; it yields a p-value of 0.0016. This statistically significant p-value suggests that a null hypothesis of no difference between the feminine oblique and the masculine oblique in the extent to which they vary would have to be rejected, and the variation in the feminine is indeed more salient in the feminine class of nouns.

It is also important to note that the vowel /e/ invariably takes part in the variation, as we see in the varying pairs of vowels: in case of the masculine oblique, the variation is /e/ ~ /o/, whereas in the feminine oblique plural it is /e/ ~ /a/. Its presence is in line with the overall high frequency of /e/, while the fact that it frequently takes part in some kind of variation is in line with the hypothesis that /e/ could be a default vowel and thus it is less stable. For instance, suffix-initial /e/ is regularly elided after a stem-final vowel in the verb morphology of Northern Vlax Romani: adding the 3rd person plural present indicative marker *-en* to the verb *asa-* ‘laugh’ yields *asán* ‘they laugh’.

6.2 Possible causes and explanations

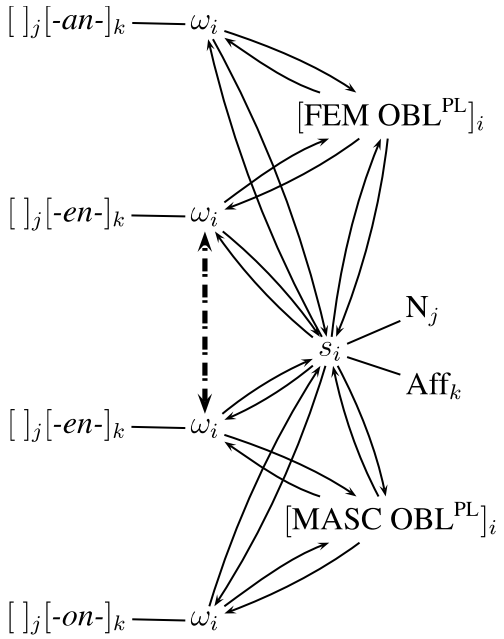
1. The masculine oblique plural *-en-*. Besides *-an-*, the other variant of the feminine oblique plural marker is *-en-*. The form is identical to one of the variants of the masculine oblique plural marker. As the semantic content (oblique plural) is also identical, we would like to look into the possible analogical influence of the masculine oblique plural marker on the feminine one. As we will see, the *-en-* form is dominant in both the masculine and the feminine nominal paradigms, which suggests that a mutual influence may exist.
2. The feminine nominative plural suffixes. We will examine whether the nominative plural endings *-i* and *-a* have any connection to the appearance of one or the other plural oblique marker. We will find that there seems to be a relationship, which is made slightly more complicated by the fact that the singular form of nouns with the plural suffix *-i* ends in *-a* and the singular form of nouns with the plural suffix *-a* often ends in *-i*.

6.2.1 *The masculine oblique plural* -en-

The presence of the *-en-* pattern in the feminine is historically related to its simultaneous presence in the masculine: the oblique plural suffix was *-en-* across the whole nominal system in the Proto-Romani stage (Elšík 2000b: 14). While the *-en-* pattern represents syncretism, exerting a neutralising effect, making all oblique plural forms look identical and decreasing the extent of the difference between the two genders, the *-an-* pattern exerts an opposite effect, favoring intra-gender uniformity, being more similar to the singular oblique marker *-a*; see for example *šūrí* ‘knife’: *šūrá-* OBL.SING ~ *šūrása* INSTR.SING and *šūrán-* OBL.PL ~ *šūránca* INSTR.PL..

The correlation between the masculine oblique plural *-en-* and the feminine oblique plural *-en-* is shown in Figure 10, where the schemas for the masculine oblique plural and the feminine oblique plural are connected through a thick dashed bidirectional arrow, indicating mutual influence. However, as we will see in Figure 11, separating the masculine and the feminine phonological components containing the *-en-* suffix is not necessary at all; just like the syntactic component, the identical phonological components can also be conflated into a single one.

Figure 10. The relationship between the masculine and the feminine oblique plural endings



Let us have a look at the phenomenon through the examples of *rakló* ‘boy’ and *rakljí* ‘girl’. These two words are apparently close cognates of each other, derived from Sanskrit *laḍikka* ‘child’ (Turner 1962-66: 633), and continuing Old Indo-Aryan masculine and feminine declension classes (Matras 2002: 42-43).

Table 11: A correlation between some masculine and feminine paradigms in Northern Vlax Romani

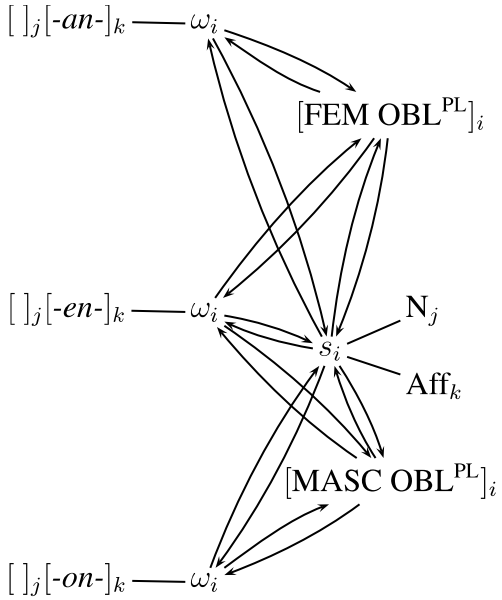
	Nominative singular	Nominative plural	Oblique singular	Oblique plural
‘boy’	*rakló*	*raklé*	*raklés-*	*raklén-*
‘girl’	*rakljí*	*rakljá*	*rakljá-*	*raklján-*

The forms in Table 11 forms show great uniformity, while maintaining opposition and differentiation. The back vowel of the nominative singular *rakló* is replaced by the front vowel /e/ in all other forms, while the front vowel of *rakljí* is replaced by the back vowel /a/ in the other forms. The back/front contrast of the nominative singular endings /o/ and /i/ is reversed in the plural and in the oblique, but the contrast between the two paradigms remains expressed. As we noted with regard to the masculine, disyllabic words always inflect the same way, having either /e/ or /o/ in the oblique ending. The word *rakló* belongs to the nouns which exhibit the *-es-/-en-* pattern.

As stated above, the overall number of masculine nouns with the *-es-/-en-* pattern is 29, as opposed to the 24 items with the *-os-/-on-* pattern (not counting the stems where there is variation). If we compare this to the seven feminine nouns with the oblique plural marker *-en-* and the four

feminine nouns with the oblique plural marker *-an-*, we can see that, at least concerning type frequency, the *-en-* form dominates in both the masculine and the feminine paradigms. The fact that there are more nouns which take the *-en-* form in both paradigms suggests that there must be a correlation between the two. Statistical evidence for the overall dominance of the *-en-* form comes from a chi-square test where we compared the proportion of nouns in *-en-* in both the masculine and the feminine paradigms, yielding a p-value of 0.5880. This means that the null hypothesis that the *-en-* form is dominant cannot be rejected:it is indeed dominant. The neutralisation effect is shown in Figure 11, where the masculine oblique plural and the feminine oblique plural converge in the ending *-en-*, and diverge through the endings *-an-* and *-on-*.

Figure 11. Combined schema of the masculine and feminine oblique plural



6.2.2 *The feminine nominative plural suffixes*

In Table 7, the feminine nominative plural form ends in /a/ if the nominative singular ends in /i/, as in *pīrí* ‘pot, saucepan’ ~ *pīrá* ‘pots, saucepans’, and the nominative plural ends in /i/ if the nominative singular ends in /a/, as in *kúrva* ‘whore’ ~ *kurví* ‘whores’. The oblique plural suffix seems to correspond to the nominative plural suffix with regard to the backness of their vowels.

(2) a. *pīrí* NOM.SING → *pīrá* NOM.PL → *pīrán-* OBL.PL
b. *kúrva* NOM.SING → *kurví* NOM.PL → *kurvén-* OBL.PL

Scrutiny of the data reveals the following numbers and proportions. Out of the total 20 items, seven items follow the vowel backness pattern: if their nominative plural ending is /i/, they will take the oblique plural suffix *-en-*, and if the nominative plural ending is /a/, they will take the oblique plural suffix *-an-*, as in Table 12.

Table 12: Feminine nouns where the nominative plural ending corresponds to the oblique plural ending in the questionnaire dataset

Nominative singular	Nominative plural	Oblique plural stem
nouns with the oblique form *-an-*		
xajíng ‘well’	*xajingá*	*xajingán-*
khajnʲí ‘hen’	*khajnʲá*	*khajnʲán-*
māčí ‘fly’	*māčá*	*māčán-*
pīrí ‘saucepan’	*pīrá*	*pīrán-*
nouns with the oblique form *-en-*		
armajá ‘curse’	*armají*	*armajén-*
kúrva ‘whore’	*kurví*	*kurvén-*
rā́ca ‘duck’	*rācí*	*rācén-*

Four items behave in the opposite way, exhibiting the nominative plural ending /a/ but the oblique plural suffix *-en-*. There are no nouns having a nominative plural in /i/ with an oblique plural stem in *-an-*.

Table 13: Feminine nouns where the nominative plural ending does not correspond to the oblique plural ending in the questionnaire dataset

Nominative singular	Nominative plural	Oblique plural stem
cincā́ri ‘mosquito’	*cincārá*	*cincārén-*
mesají ‘table’	*mesajá*	*mesajén-*
angrustí ‘ring’	*angrustá*	*angrustén-*
kangrí/krangí ‘branch’	*kangrá/krangá*	*kangrén-/krangén-*

Consider now the seven stems where there is significant variation. Three of the stems where there is variation tend to conform to the vowel backness pattern, predominantly taking either the nominative plural ending /a/ and the oblique plural suffix *-an-* or the nominative plural ending /i/ and the oblique plural suffix *-en-*.

Table 14: Feminine nouns where there is variation with a bias towards the correspondence between the nominative plural and the oblique plural in backness in the questionnaire dataset

Nominative singular	Nominative plural	Oblique plural forms		
		occurrences	in *-en-*	in *-an-*
papín ‘goose’	*papinʲá*	19	32%	68%
tʲīrí ‘ant’	*tʲīrá*	16	37.5%	62.5%
borótva ‘razor’	*borotví*	9	78%	22%

On the other hand, two of the stems with varying forms go against the tendency, with the predominant pattern being that of combining the nominative plural ending /a/ and the oblique plural suffix *-en-*.

Table 15: Feminine nouns where there is variation with a bias towards the opposition between the nominative plural and the oblique plural in backness in the questionnaire dataset

Nominative singular	Nominative plural	Oblique plural forms		
		occurrences	in *-en-*	in *-an-*
patrí ‘leaf’	*patrá*	19	79%	21%
bār ‘garden’	*bārá*	10	80%	20%

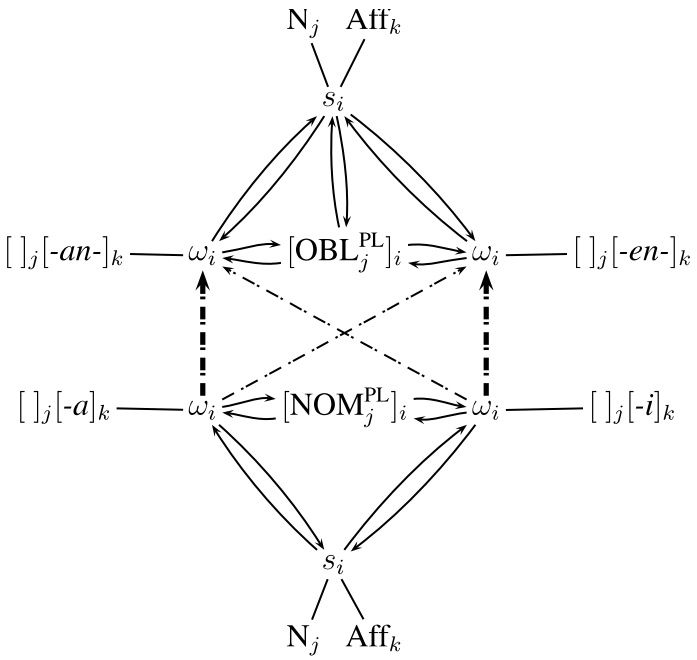
Finally, there are two stems where the proportion of the two patterns is virtually equal, indicating a high degree of variation.

Table 16: Feminine nouns where there is a considerable degree of variation with no significant bias in the questionnaire dataset

Nominative singular	Nominative plural	Oblique plural forms		
		occurrences	in *-en-*	in *-an-*
katt ‘pair of scissors’	*katt*^{j}*á*	23	48%	52%
šūrí ‘knife’	*šūrá*	25	52%	48%

In sum, we can say that the vowel backness pattern correctly predicts the oblique plural suffix from the nominative plural ending for ten stems, while this prediction goes awry in the case of only six stems. This suggests that there is a tendency for the vocalism of the feminine nominative plural ending to influence the choice of the oblique plural suffix (*-an-* if the nominative plural ending is *-a* and *-en-* if the nominative plural ending is *-i*; this tendency might be weakened by the fact that the nominative singular ending, if it is a vowel, tips the balance in favour of the other oblique plural suffix). This is shown in Figure 12, where the schemas for the nominative plural and the oblique plural are connected through dashed arrows. The thick arrows represent the dominant direction of prediction, while the thin arrows show a weak correlation.

Figure 12. The relationship between the feminine nominative plural and the feminine oblique plural as shown in the form of schemas



6.3 Interim summary

In this section, we looked at the second weak point, the feminine oblique plural base, in more detail. Following the description of the phenomenon, we examined two possible aspects that might influence the choice of the plural oblique ending for feminine nouns and we found that the two aspects indeed seem to exert influence.

1. The masculine oblique plural *-en-*. Besides *-an-*, the other variant of the feminine oblique plural marker is *-en-*, which is identical to one of the variants of the masculine oblique plural marker. We looked into the possible analogical influence of the masculine oblique plural marker on that of the feminine. As we saw, the form *-en-* is indeed dominant in both the masculine and the feminine nominal paradigms, which suggests that the influence exists.
2. The feminine nominative plural suffixes. We examined whether the nominative plural endings *-i* and *-a* have any connection to the appearance of the plural oblique marker *-en-* and *-an-*. We found that there is a relationship between the nominative and the oblique plural endings, with the front vowel /i/ predominantly predicting the marker *-en-* and the back vowel /a/ predominantly predicting the marker *-an-*.

7. General conclusion

The discussion of the two instances of variation in the nominal morphology of Northern Vlax Romani varieties as spoken in Hungary focussed on variation: the competition between different exponents expressing the same meaning. Variation in morphology has been coming under the spotlight in recent years; however, most papers have dealt with derivational and inter-dialectal variation, and less has been said on inflection and variation within a well-defined variety of a language. The present study contributes to the study of variation in inflectional morphology and attempts to demonstrate that what is commonly referred to as free variation is not entirely random or unconditional: the existence and frequency of different, competing patterns may be due to an analogy with other morphological phenomena. Variation is also conditioned in the sense that it

emerges in loci where there is a weak point in the grammar, which can be represented as an unstable correspondence in a tripartite morphological schema where all three components are interconnected.

The data clearly show that the masculine oblique base and the feminine plural oblique base in Northern Vlax Romani manifest variation. On the one hand, variation has been in abundance in Romani owing to at least two factors: firstly, it was an exclusively spoken language until recently, with numerous varieties spread over a large, often disconnected geographical area; secondly, the impact of contact with various languages throughout its history has resulted in the adoption of a range of grammatical structures by Romani as a whole and, subsequently, by its individual dialects. On the other hand, a unique feature of the language has been the fairly strict division of the lexicon into inherited and borrowed compartments, distinguished by grammatical means, among them the masculine oblique suffix *-os-*, itself borrowed from Greek; however, the interaction between the two compartments, which has already been noted, appears to have affected the oblique suffixes in several Romani dialects.

As investigated and demonstrated above, the interaction and the resulting variation may have been triggered, caused or influenced by other linguistic factors, such as the position of stress and the number of syllables (in the case of the masculine oblique base) and the masculine oblique plural *-en-* and the feminine nominative plural suffixes (in the case of the feminine oblique plural oblique base).

The Romani example described in the paper can also be seen as reflecting the simultaneous presence of two forces in language, regularisation and differentiation, which create a dynamic process within the Romani nominal system. We see two distinct but internally uniform patterns for both the masculine oblique forms and the feminine oblique plural forms. Uniformity means that we do not find mixed paradigms where the distribution of the two patterns would show regularity for one particular noun but it also refers to the presence of the marker *-en-* in the feminine plural oblique, reflecting variation within the feminine plural paradigms but uniformity in the wider category of nouns.

Acknowledgements

I would like to thank Jenny Audring, László Kálmán, Mikós Törkenczy and the anonymous reviewers for their invaluable comments on previous versions of this paper.
The support of the National Research, Development and Innovation Fund (grants K 111961 and K 125596, project leader: László Kálmán) is gratefully acknowledged.

References

Adamou, Evangelia. 2016. *A corpus-driven approach to language contact.* Berlin: Mouton de Gruyter.

Austin, Jennifer. 2012. Markedness, input frequency, and the acquisition of inflection: Evidence from Basque/Spanish bilingual children. *International Journal of Bilingualism* 17(3). 259–283.

Bakker, Peter. 1997. Athematic morphology in Romani: The borrowing of a borrowing pattern. In Matras, Yaron & Bakker, Peter & Kyuchukov, Hristo (eds.), *The typology and dialectology of Romani*, 1–21. Amsterdam: John Benjamins.

Bakker, Peter. 2001. Romani in Europe. In Extra, Guus & Gorter, Durk (eds.), *The Other Languages of Europe*, 293–313. Clevedon: Multilingual Matters.

Baló, Márton A. 2008. A strange case of defectiveness in the Lovari verbal paradigm. In László, Kálmán (ed.), *Papers from the Mókus Conference*, 118–136. Budapest: Tinta.

Barbiers, Sjef & Cornips, Leonie. 2007. Introduction to syntactic microvariation. In Barbiers, Sjef & Cornips, Leonie & van der Kleij, Susanne (eds.), *Syntactic microvariation*, 2–12. Amsterdam: Meertens Institute.

Barbiers, Sjef & Cornips, Leonie & van der Kleij, Susanne (eds.). 2007. *Syntactic microvariation*. Amsterdam: Meertens Institute.

Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories.* Berlin: Mouton de Gruyter.

Blything, Ryan P. & Ambridge, Ben & Lieven, Elena V. M. 2018. Children's acquisition of the English past tense: Evidence for a single-route account from novel verb production data. *Cognitive Science* 42. 621–639.

Booij, Geert. 2010. *Construction Morphology.* Oxford: Oxford University Press.

Boretzky, Norbert. 1989. Zum Interferenzverhalten des Romani (Verbreitete und ungewöhnliche Phänomene). *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42. 357–374.

Boretzky, Norbert. 1994. *Romani: Grammatik des Kalderaš-Dialekts mit Texten und Glossar.* Wiesbaden: Harrassowitz.

Boretzky, Norbert. 2003. *Die Vlach-Dialekte des Romani.* Wiesbaden: Harrassowitz.

Boretzky, Norbert & Igla, Birgit. 2004. *Kommentierter Dialektatlas des Romani*. Wiebaden: Harrassowitz.

Brandner, Ellen. 2012. Syntactic microvariation. *Language and Linguistics Compass* 6. 113–130.

Bybee, Joan L. 2013. Usage-based theory and exemplar representations of constructions. In Hoffmann, Thomas & Trousdale, Graeme (eds.), *The Oxford handbook of Construction Grammar*, 49–69. Oxford: Oxford University Press.

Bybee, Joan L. & Moder, Carol L. 1983. Morphological classes as natural categories. *Language* 59. 251–270.

Cech, Petra & Fennesz-Juhasz, Christiane & Heinschink, Moses F. 1999. *Texte österreischer Lovara I. Arbeitsbericht 2 des Projekts Kodifizierung der Romanes-Variante der österreichischen Lovara (hrsgg. von Dieter W. Halwachs)*. Vienna: Verein Romano Centro.

Cech, Petra & Heinschink, Moses F. 1999. *Basisgrammatik. Arbeitsbericht 1a des Projekts Kodifizierung der Romanes-Variante der österreichischen Lovara (hrsgg. von Dieter W. Halwachs).* Vienna: Verein Romano Centro.

Corbett, Greville G. 2009. Canonical inflectional classes. In Montermini, Fabio & Boyé, Gilles & Tseng, Jesse (eds.), *Selected proceedings of the 6th Décembrettes*, 1–11. Somerville, MA: Cascadilla Proceedings Project.

Dammel, Antje & Schallert, Oliver (eds.). 2019a. *Morphological variation: Theoretical and empirical perspectives*. Amsterdam: John Benjamins.

Dammel, Antje & Schallert, Oliver. 2019b. Introduction. In Dammel, Antje & Schallert, Oliver (eds.), *Moprhological variation: Theoretical and empirical perspectives*. 1–25. Amsterdam: John Benjamins.

Deuchar, Margaret. 2006. Welsh-English code-switching and the Matrix Language frame model. *Lingua* 116(11). 1986–2011.

Dufter, Andreas & Fleischer, Jürg & Seiler, Guido. 2009a. Introduction. In Dufter, Andreas & Fleischer, Jürg & Seiler, Guido (eds.), *Describing and modeling variation in grammar.* 1–18. Berlin: Mouton de Gruyter.

Dufter, Andreas & Fleischer, Jürg & Seiler, Guido (eds.). 2009b. *Describing and modeling variation in grammar.* Berlin: Mouton de Gruyter.

Elšík, Viktor. 2000a. Dialect variation in Romani personal pronouns. In Elšík, Viktor & Matras, Yaron (eds.), *Grammatical relations in Romani: The noun phrase*, 65–94. Amsterdam: John Benjamins.

Elšík, Viktor. 2000b. Romani nominal paradigms: Their structure, diversity, and development. In Elšík, Viktor & Matras, Yaron (eds.), *Grammatical relations in Romani: The noun phrase*, 9–30. Amsterdam: John Benjamins.

Elšík, Viktor & Matras, Yaron. 2006. *Markedness and language change: The Romani sample.* The Hague: Mouton de Gruyter.

Erdős, Kamill. 1959. *A Békés megyei cigányok. Cigánydialektusok Magyarországon*. Gyula: Erkel Ferenc Múzeum.

Gaspar, Lúcia. 2012. Ciganos no Brasil. *Pesquisa Escolar Online.* Recife: Joaquim Nabuco Foundation. (http://basilio.fundaj.gov.br/pesquisaescolar_en/index.php?option=com_content&id=1263:gypsies-in-brazil-ciganos-no-brasil.) (Accessed June 2020)
Gilliat-Smith, Bernard J. 1915. A report on the Gypsy tribes of North East Bulgaria. *Journal of the Gypsy Lore Society, new series* 9. 1–54, 65–109.
Goldberg, Adele E. 1995. *Constructions.* Chicago: The University of Chicago Press.
Goldberg, Adele E. 2013. Argument structure constructions vs. lexical rules or derivational verb templates. *Mind and Language* 28(4). 436–465.
Hancock, Ian F. 2013. The schooling of Romani Americans: An overview. In Miskovic, Maja (ed.), *Roma education in Europe: Practices, policies and politics*, 87–100. New York, NY: Routledge.
Harris, Alice C. 2017. *Multiple exponence.* Oxford: Oxford University Presss.
Hutterer, Miklós & Mészáros, György. 1967. *A lovári cigány dialektus leíró nyelvtana.* Budapest: Magyar Nyelvtudományi Társaság.
Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41. 9–45.
Jackendoff, Ray. 2008. *Construction after construction* and its theoretical challenges. *Language* 84: 8–28.
Jackendoff, Ray. 2012. Language. In Frankish, Keith & Ramsey, William M. (eds.), *The Cambridge handbook of cognitive science*, 171–193. Cambridge: Cambridge University Press.
Jackendoff, Ray & Audring, Jenny. 2016. Morphological schemas. *The Mental Lexicon* 11(3). 467–493.
Kálmán, László & Rebrus, Péter & Törkenczy, Miklós. 2012. Possible and impossible variation in Hungarian. In Kiefer, Ferenc & Siptár, Péter & Ladányi, Mária (eds.), *Current issues in morphological theory: (Ir)regularity, analogy and frequency*, 23–49. Amsterdam: John Benjamins.
Kozhanov, Kirill. 2016. Is there a “new infinitive” in Russian Romani? A corpus-based study of subject-verb agreement in the subjunctive. Paper presented at the Conference on Corpus-Based Approaches to the Balkan Languages and Dialects, St Petersburg, Russia, 5–7 December 2016.
Leggio, Daniele Viktor & Matras, Yaron. 2017. Variation and dialect levelling in the Romani dialect of Tăndărei. *Romani Studies 5* 27(2). 173–209.
Matras, Yaron. 2002. *Romani: A linguistic introduction.* Cambridge: Cambridge University Press.
Matras, Yaron. 2005a. The status of Romani in Europe. Report submitted to the Council of Europe’s Language Policy Division. (http://romani.humanities.manchester.ac.uk/downloads/1/statusofromani.pdf) (Accessed December 2013.)
Matras, Yaron. 2005b. The classification of Romani dialects: A geographic-historical perspective. In Schrammel, Barbara & Halwachs Dieter W. (eds.), *General and applied Romani linguistics*, 7–26. Munich: Lincom Europa.
Matras, Yaron & Adamou, Evangelia. 2020. Romani and contact linguistics. In Matras, Yaron & Tenser, Anton (eds.), *The Palgrave handbook of Romani linguistics*, 329–352. London: Palgrave Macmillan.
Miklosich, Franz. 1872-80. *Über die Mundarten und Wanderungen der Zigeuner Europas*. Vienna: Karl Gerold’s Sohn.
Neef, Martin. 2009. Morphological variation: A declarative approach. In Dufter, Andreas & Fleischer, Jürg & Seiler, Guido (eds.), *Describing and modeling variation in grammar*, 117–133. Berlin: Mouton de Gruyter.
Paster, Mary. 2010. Optional multiple plural marking in Maay. In Rainer, Franz & Dressler, Wolfgang U. & Kastovsky, Dieter & Luschützky, Hans Christian. (eds.), *Variation and change in morphology. Selected papers from the 13th International Morphology Meeting, Vienna, February 2008*, 177–192. Amsterdam: John Benjamins.

Pierrehumbert, Janet P. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, Joan & Hopper, Paul (eds.), *Frequency and the emergence of linguistic structure*, 139–157. Amsterdam: John Benjamins.

Poplack, Shana. 1993. Variation theory and language contact. In Preston, Dennis (ed.), *American dialect research*, 251–286. Amsterdam: John Benjamins.

Rainer, Franz & Dressler, Wolfgang U. & Kastovsky, Dieter & Luschützky, Hans Christian. 2010a. Introduction. In *Variation and change in morphology. Selected papers from the 13th International Morphology Meeting, Vienna, February 2008.* 1–13. Amsterdam: John Benjamins.

Rainer, Franz & Dressler, Wolfgang U. & Kastovsky, Dieter & Luschützky, Hans Christian (eds.). 2010b. *Variation and change in morphology. Selected papers from the 13th International Morphology Meeting, Vienna, February 2008.* Amsterdam: John Benjamins.

Rebrus, Péter & Törkenczy, Miklós. 2011. Paradigmatic variation in Hungarian. In Laczkó, Tibor & Ringen, Catherine F. (eds.), *Approaches to Hungarian, Volume 12: Papers from the 2009 Debrecen Conference,* 135–162. Amsterdam: John Benjamins.

Rumelhart, David E. 1980. Schemata: The building blocks of cognition. In Spiro, Rand J. & Bruce, Bertram C. & Brewer, William F. (eds.), *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education*, 33–58. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sloetjes, Han, & Wittenburg, Peter. 2008. Annotation by category – ELAN and ISO DCR. In Calzolari, Nicoletta & Choukri, Khalid & Maegaard, Bente & Mariani, Joseph & Odijk, Jan & Piperidis, Stelios & Tapias, Daniel (eds.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Paris: European Language Resources Association.

Tálos, Endre. 2001. A cigány és a beás nyelv Magyarországon. In Kovalcsik, Katalin & Csongor, Anna & Bódi, Zsuzsanna (eds.), *Tanulmányok a cigányság társadalmi helyzete és kultúrája köréből,* 317–324. Budapest: Eötvös Loránd Tudományegyetem – Iskolafejlesztési Alapítvány – Oktatási Minisztérium.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Turner, Ralph L., Sir. 1962-1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

van den Heuvel, Wilco & Urech, Evelyne. 2014. Romani dialect variation in Transylvania: Migration and diffusion. *Romani Studies 5* 24(1). 43–70.

Vekerdi, József. 1985. *Cigány nyelvjárási népmesék.* Debrecen: Kossuth Lajos Tudományegyetem.

Vekerdi, József. 2000. *A comparative dictionary of Gypsy dialects in Hungary*. Budapest: Terebess.