Romani nominal paradigms: their structure, diversity and development

Viktor Elšík

Manuscript 1999, published as:
ROMANI NOMINAL PARADIGMS
THEIR STRUCTURE, DIVERSITY, AND DEVELOPMENT

VIKTOR ELŠÍK
Charles University, Prague & University of Manchester

1. Introduction

In the present contribution I intend to consider paradigm diversity of Romani nouns and adjectives. Two dimensions of paradigm diversity are taken into account: the cross-dialectal diversity of genetically, formally, or functionally related paradigms, and the diversity of paradigms within a dialect. The dialectological perspective, of course, brings the need for a diachronic approach. However, I do not attempt a full historical description of all paradigms in all dialects. Instead, I employ a model of four stages of development of Romani inflectional paradigms (A, B, C, and D), which enables one to have a general insight into the fundamental drifts in the development of the classification morphology of Romani nouns and adjectives.

Stage A (Section 3) represents the genetic heritage from Indo-Aryan in Proto-Romani, the pre-European ancestor of the Romani language. I will briefly discuss Proto-Romani inherited classification and some later developments within the inherited classes. In Stage B (Section 4) a new, contact-induced pattern was introduced into Romani with far-reaching consequences for paradigm structure, morphemic segmentation, and classification. The central notion in any discussion of the new pattern is the concept of athematic morphology. Stage B developments came about at the time of a relative linguistic unity of Romani (termed Early Romani) in the Greek linguistic area. Stages C and D, respectively, represent the ‘life’ and the decline of Stage B pattern in individual Romani dialects or dialect groups. While the general processes characteristic of Stage C (Section 5) are common to all Romani dialects, Stage D (Section 6), which is manifested by class extinctions and blurring of the thematicity dichotomy, has been reached only in some of them.
Before proceeding to the Stage-related phenomena, a few problematic issues concerning the general structure of Romani noun paradigms will be addressed in the following section.

2. The structure of Romani nominal paradigms

2.1 Subject vs. direct object split

There is an inherited split in encoding the subject – direct object relation in Romani. Direct object marking may be either distinct from subject marking (with nominals higher on an animacy scale), or homonymous with it (with nouns lower on the scale). The traditional analysis is presented in Figure 1:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Direct object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animate</td>
<td>Inanimate</td>
</tr>
<tr>
<td>dad “father” = NOM</td>
<td>kher “house” = NOM</td>
</tr>
<tr>
<td>dades = ACC</td>
<td>kher = ACC</td>
</tr>
</tbody>
</table>

Figure 1: Case marking of direct object: traditional analysis

The form kher “house” in the direct object function has been traditionally analyzed as an accusative form homonymous with the nominative form. This inevitably results in describing the analogical inflection of animate and inanimate nouns as separate paradigms.

I propose instead to consider the direct object form kher to be a nominative. Consider the following examples:

(1) East Slovak Romani
   a. mir -i rokl’ -a puraň -i
      my -NOM.SG.F skirt -NOM.SG old   -NOM.SG.F
      “my skirt is old”
   b. dikhes mir -i rokl’ -a?
      you.see my -NOM.SG.F skirt -NOM.SG
      “can you see my skirt?”

(2) a. mir -i kamaratk -a terň -i
     my -NOM.SG.F female.friend -NOM.SG young -NOM.SG.F
     “my friend is young”
   b. dikhes mir -a kamaratk -a
      you.see my -OBL.SG.F female.friend -ACC.SG
      “can you see my friend?”
The NPs in examples (a) are in the subject position, while the NPs in examples (b) are direct objects. Both head nouns of these NPs – the inanimate rokla “skirt” in (1) as well as the animate kamaratka “female friend” in (2) – exhibit identical forms in both constructions. In (1) the whole object NP is identical to the subject NP. In (2), however, the differing agreement marking on the possessive pronoun mir- “my” shows that the head noun takes a different case in each syntactic function. It is obvious that in case of animate nouns such as kamaratka “female friend” we are dealing with a morphological nominative-accusative homonymy. In the case of inanimate nouns, on the other hand, the nominative rather than the accusative is used to mark the direct object. This analysis is presented in Figure 2:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Direct object</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dad</em> “father” = NOM</td>
<td><em>dades</em> = ACC</td>
</tr>
<tr>
<td><em>kher</em> “house” = NOM</td>
<td><em>(kheres)</em> = ACC</td>
</tr>
</tbody>
</table>

*Figure 2: Case marking of direct object*

Although the ‘real’ accusative forms of inanimate nouns (such as *kheres* in Figure 2) are not common in the language, their analogical creation may be triggered by language interference. Consider the next examples:

(3)  

a. East Slovak Romani  

```
le rom -es hin bang -e  
the.OBL.SG.M man -ACC.SG COP crooked -NOM.PL  
čang -a  
leg -NOM.PL  
“the man has crooked legs”
```

b.  

```
o kher hin bang -e  
the.NOM.SG.M house COP crooked -OBL.PL  
fal -en -ger -o  
wall -OBL.PL -GEN -NOM.SG.M  
“the house has crooked walls”
```

c.  

```
le kher -es hin bang -e  
ART.OBL.SG.M house -ACC.SG COP crooked -NOM.PL  
fal -i  
wall -NOM.PL
```
“the house has crooked walls”

In (3a) we have a regular East Slovak Romani possessive construction with an animate possessor in the accusative and a possessee in the nominative. In most varieties the possessive construction is not common with inanimate possessors. Instead, other constructions are employed, such as the one in (3b). Nevertheless, Slovak, the main contact language of this dialect, does not distinguish animate and inanimate possessors, using a possessive construction with both. The influence of Slovak then leads to sentences such as (3c) in some varieties or idiolects. Romani calques the distribution of the possessive construction, and the construction itself then triggers the use of the accusative with inanimate nouns as well.

To conclude this section: It is more adequate to consider the pairs of nouns such as *dad* “father” and *kher* “house” (see Figures 1 & 2) as belonging to one and the same paradigm, i.e. to treat the subject – direct object split as a (hyper-paradigmatic) phenomenon which does not contribute to the complementary classification of nouns.

2.2 *Oblique-stem agglutination*

Many Romani noun forms, namely the oblique cases, are clearly agglutinative in their structure (cf. Friedman 1991, Elšík 1997). Nevertheless, Romani nominal paradigms differ in some respects from the type of agglutinative declension found, for example, in Quechua, or most Altaic and Uralic languages. I propose to distinguish between base-stem agglutination (BSA) which is present in these languages, and oblique-stem agglutination (OSA), which exists, for example, in Daghestanian languages and – also in Romani. The basic difference concerns marking of the category of case. In BSA, cases are expressed by individual suffixes, while the case suffixes in OSA combine with an oblique formant.

At the same time the oblique formants are sensitive to classification, both in Romani and the Daghestanian languages (cf. Kibrik 1991). More generally, classification of Romani nouns is located in the so-called Layer I segments (for the term cf. Matras 1997), and in stem formatives such as jotation (see 3.1). In the course of time, Romani case suffixes have been becoming less and less agglutinative, losing their surface uniformity due to phonological processes. Layer I segments, by contrast, have acquired a more agglutinative structure through morphological re-segmentation, which was brought about by Stage B contact-induced developments (see 4.2). In OSA one of the oblique cases is
usually unmarked with respect to the others. It is the ergative in most Daghestanian languages, and the accusative in Romani (see 2.3).

2.3 Accusative vs. oblique stem

There is no general agreement regarding the paradigmatic status of Romani accusative forms. According to some authors (e.g. Hübschmannová et al. 1991: 617) the accusative form provides (in synchronic terms) a base for the formation of the other oblique cases. I propose, instead, that the accusative form itself is derived from the same morphological construction as the other oblique cases – viz. from the oblique stem – by an identity process (OBL ø → ACC). The former may be termed accusative analysis, the latter oblique-stem analysis. Following Kibrik’s (1991:257) argumentation on the status of the ergative within noun paradigms in Daghestanian languages, I maintain that the accusative analysis must be refused for semantic reasons: there is no accusative component in the meaning of the locative, instrumental etc. The advantage of the oblique-stem analysis, on the other hand, is that it does not need to resort to deriving an inflectional form from another inflectional form, still being able to award the accusative an unmarked status as against the other oblique cases.

Nevertheless, there are instances in a few Romani dialects of a systematic split between the surface forms of the accusative and the oblique stem. Some of these are of a phonotactic rather than morphological nature. The accusative forms may be arrived at by applying phonotactic rules. For example, in Arli the accusative dad-e is derived from the oblique dad-es- by an identity derivation plus a phonotactic rule which does not permit word-final s in grammatical formants (s > ø l –#). In the Southern Central dialects, unlike Arli, final s-lessness is not automatic in grammatical formants. Nevertheless, the split between the oblique (e.g. dad-es-) and the s-less accusative (e.g. dad-e) can still be reconciled by a less straightforward sort of oblique stem analysis. Such an analysis would derive the accusative from the oblique through morphological deletion (OBL -s→ ACC).

Morphological deletion would also account for a number of instances in Roman (e.g. OBL dad-es−→ ACC dad-e). With some Roman nouns, however, the accusative form must be derived from the base stem (e.g. grof−→ ACC grof-e), being quite independent of the oblique (e.g. grof-os-). Thus in Roman the split between the accusative and the oblique concerns not only their surface forms but also their morphological structure.
3. **Stage A: Proto-Romani inheritance**

3.1 **Nouns**

Noun classes as reconstructed for Proto-Romani are given in Figure 3:

<table>
<thead>
<tr>
<th>Class</th>
<th>Symbol</th>
<th>Example</th>
<th>NOM SG</th>
<th>NOM PL</th>
<th>OBL SG</th>
<th>OBL PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>zero-masculines</td>
<td>MØ-a</td>
<td>kher “house”</td>
<td>-</td>
<td>-a</td>
<td>-es-</td>
<td>-en-</td>
</tr>
<tr>
<td></td>
<td>MØ-A</td>
<td>čačipen “truth”</td>
<td>-</td>
<td>-a</td>
<td>-as-</td>
<td>-en-</td>
</tr>
<tr>
<td></td>
<td>MØ-Ø</td>
<td>vast “hand”</td>
<td>-</td>
<td>–</td>
<td>-es-</td>
<td>-en-</td>
</tr>
<tr>
<td>o-masculines</td>
<td>Mo</td>
<td>šero “head”</td>
<td>-o</td>
<td>-e</td>
<td>-es-</td>
<td>-en-</td>
</tr>
<tr>
<td>i-masculines</td>
<td>Mi</td>
<td>pani “water”</td>
<td>-i</td>
<td>-j-a</td>
<td>-j-es</td>
<td>-j-en</td>
</tr>
<tr>
<td>zero-feminines</td>
<td>FØ-U</td>
<td>džuv “louse”</td>
<td>-j-a</td>
<td>-a</td>
<td>-a</td>
<td>-en-</td>
</tr>
<tr>
<td></td>
<td>FØ-J</td>
<td>suv “needle”</td>
<td>-j-a</td>
<td>-j-a</td>
<td>-j-a</td>
<td>-j-en</td>
</tr>
<tr>
<td>i-feminines</td>
<td>Fi</td>
<td>piri “pot”</td>
<td>-i</td>
<td>-j-a</td>
<td>-j-a</td>
<td>-j-en</td>
</tr>
</tbody>
</table>

*Figure 3: Proto-Romani noun classes*

The class symbols in the second column of the Figure render the criteria of Romani noun classification. Two criteria are general for all nouns: the gender, and the shape of the base-form (i.e. NOM.SG) suffix. Thus, for example, the zero-feminines (FØ) are feminine nouns whose NOM.SG form does not show a positive inflectional suffix – in practice, the base form ends in a consonant. The two basic criteria classify nouns into five groups: zero-masculines, o-masculines, i-masculines, zero-feminines, and i-feminines. However, not all members of a zero-class (of a certain gender) inflect alike.

Three subclasses must be distinguished within Proto-Romani zero-masculines. Abstract nouns (MØ-A), which are characterized by a specific derivational suffix, possess a different OBL.SG suffix (-as-) than the other two subclasses. These in turn differ from each other in the NOM.PL suffix, which is either an -a (MØ-a) or a zero (MØ-Ø). Within zero-feminines two subclasses are distinguished according to the presence or absence of jotation as a stem modification in non-base forms, i.e. in all forms with the exception of the NOM.SG. The subclass with jotation may be termed *jotated* zero-feminines (FØ-J), as against the *unmodified* zero-feminines (FØ-U).

Figure 3 contains an overview of the nominative and oblique formants of the individual noun classes. It may be observed that jotation is present not only in jotated zero-feminines, but also in the i-subclasses of both genders. The formants of the non-base forms in these subclasses thus consist of two segments: stem-modifying jotation, and a categorial marker. The OBL.PL marker -en- is
general for all subclasses, i.e. selected by default. The OBL.SG marker is predictable from gender: feminines have -a, while masculines have -es- ~ -as- (as has been mentioned, the distribution of both allomorphs is predictable from derivational structure of the noun). With the ‘vocalic’ (i.e. o- and i-) classes also the NOM.PL marker and the presence or absence of jotation are predictable. They can be inferred from the shape of the base form. Thus o-nouns will have the NOM.PL suffix -e and no jotation, while i-nouns (irrespective of gender) will show positive jotation and NOM.PL -a. On the other hand, the NOM.PL formant is not predictable for zero-masculines, and neither the NOM.PL formant nor the jotation are predictable for zero-feminines.

The implicational structure of Proto-Romani noun paradigms enables us to explain some subsequent developments within the inherited classes, for example, the de-jotation of the zero-feminines. The de-jotation (e.g. *suv-j-a > suv-a “needles”) has been accomplished independently in a number of dialects (e.g. in Ajia Varvara, the Northern Central dialects, and some Lovari), so it seems there must be a ‘natural’ reason for this development. The argument could be as follows: The membership of consonant-final feminines in one of the two zero-feminine subclasses, the jotated and the unmodified, was not predictable from any kind of extramorphological data (such as shape of the base form, gender, or semantics of the noun). In fact it was the jotation which was unpredictable, so there had to be another reference form stored in the lexicon for the jotated zero-feminines. According to Wurzel (1990), classes with marked lexical representation are ‘unstable’ and easily lost – and this is exactly what has happened in some dialects with the jotated zero-feminines.

Moreover, the class of jotated zero-feminines may be analyzed as a mixed class, having its NOM.SG formant in common with unmodified zero-feminines, and the other formants with i-feminines:

<table>
<thead>
<tr>
<th></th>
<th>NOM.SG</th>
<th>NOM.PL</th>
<th>OBL.SG</th>
<th>OBL.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>FØ-U</td>
<td>-</td>
<td>-a</td>
<td>-a</td>
<td>-en-</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FØ-J</td>
<td>-</td>
<td>-j-a</td>
<td>-j-a</td>
<td>-j-en-</td>
</tr>
<tr>
<td></td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Fi</td>
<td>-i</td>
<td>-j-a</td>
<td>-j-a</td>
<td>-j-en-</td>
</tr>
</tbody>
</table>

*Figure 4: Jotated zero-feminines as a mixed class*
According to Carstairs-McCarthy’s (1991) paradigm economy principle, mixed
classes should not exist, or, if they do, they are likely to be lost. Thus it seems
that the pre-conditions for this de-jotation had been inherited, and present
already in Proto-Romani.

3.2 Adjectives

There were two adjective classes in Proto-Romani: the declinable and the
indeclinable adjectives. The inflection of the former are given in the next
figure:

<table>
<thead>
<tr>
<th>NOM</th>
<th>SG.M</th>
<th>SG.F</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBL</td>
<td>SG.M</td>
<td>SG.F</td>
<td>PL</td>
</tr>
<tr>
<td>-o</td>
<td>-i</td>
<td>-e</td>
<td>-e</td>
</tr>
<tr>
<td>-a</td>
<td>-e</td>
<td>-e</td>
<td>-e</td>
</tr>
</tbody>
</table>

Figure 5: Proto-Romani declinable adjectives

The Proto-Romani OBL.SG.F suffix -α has been retained in a few dialects (e.g.
Prizren, all Central, most Lovari, and Taikon Kalderaš). In most dialects, how-
ever, the original OBL.SG.F suffix has been replaced by an -e, which lead to a
complete neutralization of the number/gender distinctions in the oblique.

The homonymy pattern in the masculine subparadigm of the declinable
adjectives in Romani is also the most common one in a majority of the
adjective-inflecting Indo-Aryan languages (e.g. in Gujarati, Rajasthani, Hindi,
or Parya). The NOM.SG form is distinct from the other forms, which are all
homonymous (e.g. bar-o vs. bar-e “big”, respectively).\(^4\) Two homonymy
patterns exist in the feminine subparadigm, according to the dialect in question.
There are either three distinct forms (e.g. NOM.SG bar-i vs. OBL.SG bar-a vs.
PL bar-e), or two (e.g. NOM.SG bar-i vs. OBL.SG/PL bar-e). In any case, the
Romani feminine homonymy pattern is quite specific among the NIA
languages.\(^5\) The reason is that Romani – unlike any Subcontinental NIA
language, but like Domari (cf. Macalister 1910) – has undergone a masculine
take-over in the plural, which (as a type of systematic homonymy) has been
retained even in the later created athematic classes.
4. **Stage B**

4.1 **Marked integration**

It is a well-established fact that nouns borrowed into Romani before a certain date in the history of the language had in general been integrated into the inherited inflection classes. Since such borrowings are morphologically indistinguishable from the Indo-Aryan nouns, one may speak of their unmarked integration. The stratum of unmarkedly integrated loans has been conveniently termed Asian (cf. Bakker 1997:2). The unmarked integration was also common with early borrowings from Greek, e.g. *drom* “road”. It was the inflectional stem rather than the base form which was adopted:

<table>
<thead>
<tr>
<th>Greek</th>
<th>NOM.SG</th>
<th>NOM.PL</th>
<th>OBL.SG</th>
<th>OBL.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>drom-os</td>
<td>drom-i</td>
<td>[drom-]</td>
<td>[drom-]</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 6: Unmarked integration of Greek nouns*

After a certain period of intense Greek influence, however, new borrowings started to behave quite differently. Consider the Early Romani inflection of the noun *foros* “town”:

<table>
<thead>
<tr>
<th>Greek</th>
<th>NOM.SG</th>
<th>NOM.PL</th>
<th>OBL.SG</th>
<th>OBL.PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>for-os</td>
<td>for-i</td>
<td>[for-]</td>
<td>[for-]</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 7: Marked integration of Greek nouns*

Basically, it was the inflectional stem *for-* which was adopted, but some forms – viz. nominatives, as one may expect, taking into account their highest frequency and categorial unmarkedness – have been borrowed as wholes, together with their Greek inflections. Nevertheless, the borrowed forms were still analyzable into a stem and an inflection in Romani, thus giving rise to new inflectional suffixes (e.g. NOM.SG -os, NOM.PL -i), and, consequently, to new, athematic, inflection classes.

The most important fact is that once a noun was integrated into an athematic class, it was immediately marked as a borrowing by its very membership in such a class. Marked integration and marked inflection classes by itself are nothing exceptional in the world’s languages. The curious thing in
Romani is that the newly arisen classes had not remained closed and limited to their constituting, i.e. Greek, lexical stratum. On the contrary, the athematic classes have become the only ones which exhibit any degree of contact-productivity. Basically all post-Greek noun loans have been integrated into the new, athematic, rather than the old, thematic, classes. At the same time, the thematic classes retain a certain degree of internal productivity, since newly derived nouns (admittedly low in number) mostly inflect as thematics.7

4.2 Athematic classes

The Early Romani athematic classes are partly reconstructed in the following Figure (the athematic classes are as a rule symbolized by an asterix):

<table>
<thead>
<tr>
<th>Class</th>
<th>Symbol</th>
<th>Example</th>
<th>NOM SG</th>
<th>PL</th>
<th>OBL SG</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-masculines</td>
<td>*Mo</td>
<td>foros “town”</td>
<td>-o(s)</td>
<td>-i</td>
<td>-os</td>
<td>-en</td>
</tr>
<tr>
<td>u-masculines 8</td>
<td>*Mu</td>
<td>papus “grandfather”</td>
<td>-u(s)</td>
<td>-i</td>
<td>-us</td>
<td>-en</td>
</tr>
<tr>
<td>i-masculines</td>
<td>*Mi</td>
<td>sapunis “soap”</td>
<td>-(s)</td>
<td>-ja</td>
<td>-es-?</td>
<td>-en</td>
</tr>
<tr>
<td>a-feminines</td>
<td>*Fa</td>
<td>cipa “skin”</td>
<td>-a</td>
<td>?</td>
<td>-a</td>
<td>-en</td>
</tr>
</tbody>
</table>

*Figure 8: Athematic noun classes in Early Romani*

The athematic OBL.PL suffix was identical with the thematic one. An intraparadigmatic levelling must be assumed in the OBL.SG forms of the athematic masculines, still in Early Romani, for one to be able to arrive at the current forms (e.g. *for-es → for-os, *pap-es → pap-us in analogy to the NOM.SG). It is likely that the OBL.SG -is- of the athematic i-masculines, which has developed by the same type of analogy in a number of dialects but which does not exist in many others, is a later development.

An important consequence of the existence of the athematic classes was a re-analysis of the OBL.SG formant in masculines, which broke the inflective character of the masculine Layer I formants (e.g. -es- → -e-s-). The formant now consisted of the OBL.SG.M marker (-s-) and a classification vowel (-e- ~ -a-, -o-, -u-, and in some dialects -i-):

<table>
<thead>
<tr>
<th>OBL.SG</th>
<th>Mo &amp; MØ</th>
<th>MØ-A</th>
<th>*Mo</th>
<th>*Mu</th>
<th>*Mi</th>
</tr>
</thead>
</table>
| -e-s-  | -a-s-  | -o-s-| -s- | -a-s-| -i-s-  (-e-s-)

*Figure 9: Bimorphemic structure of OBL.SG.M formants*
The new type of classification with a specific classification marker has the character of a general pattern, to which new members may be added. In some dialects, for example, masculines ending in a have been adapted in such a way that they created the class of athematic a-masculines (cf. Romungro NOM.SG komunišť-a “communist”, OBL.SG komunišť-a-s-, a word borrowed from Hungarian). In a similar vain some Balkan dialects (e.g. Bugurdži) have developed classes with a stressed classification vowel (symbolized as *MV, see also 5.4).

5. **Stage C**

5.1 **Gender and inflection class assignment**

Since the Greek period Romani has been borrowing nouns from languages with various gender systems. It is a well known fact that gender and inflection class need not coincide: not only may there be more than one inflection class per gender, but also nouns of differing genders may have an identical inflection. The latter is almost never the case in Romani: there are only exclusively masculine and exclusively feminine noun classes. Criteria for assigning gender and class to a loan into Romani are: a) natural sex of the referent of the borrowed noun – only with humans and high animates; b) grammatical gender of the noun in the source language, if any; c) gender of the noun which is being replaced by the borrowing, if this is the case; and d) phonological shape of the noun in the source language – basically, the final phonemes, and sometimes the number of syllables (cf. Elšík et al. 1999:320-327).

Natural sex determines gender in Romani (i.e. male → masculine, female → feminine), and this criterion operates almost without exceptions. The only exception I know of is from Bugurdži:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Gender</th>
<th>Inflection class</th>
<th>ART agreement</th>
<th>ADJ agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>mamos</em> “granny”</td>
<td>female</td>
<td><em>Mo</em></td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td><em>sluga</em> “servant”</td>
<td>male</td>
<td><em>Fa</em></td>
<td>F</td>
<td>M</td>
</tr>
</tbody>
</table>

**Figure 10: Anti-semantic class assignment in Bugurdži**

The two nouns in Figure 10 are assigned gender according to formal criteria, which exceptionally override the semantic criterion. (There are also some peculiarities in agreement, which need not concern us here.)
Masculine or feminine gender of a borrowed noun is usually retained in Romani, but there are also a number of exceptions (e.g. masculine vecos “matter” in East Slovak Romani from the Slovak feminine vec, or feminine berga “mountain” in Northeastern dialects from the German masculine Berg). Almost regular feminine assignment occurs with names of localities, at least in the Northern (e.g. Xaladytko Minska from the masculine Minsk, Sinti Frankfurt from the neuter Frankfurt) and the Central dialects (e.g. Zohra from the Slovak masculine Zohor, Hradišča from the neuter Hradište, Šmiža from the plurale tantum Smižany, or Šoka from the genderless Hungarian Sók).

The most problematic seems to be the gender assignment to source neuters: Slavic neuters usually become masculines (e.g. Bohemian Romani caklos “glass” from Serbian/Croatian caklo, mořos “sea” from Czech moře, Romungro nebo “sky, heaven” from Serbian/Croatian nebo, and nebespečo “danger” from Slovak dialectal nebezpečí, or Bugurdžić bremes “burden” from Serbian breme). German neuters, however, are mostly assigned the feminine gender in Sinti and some Northern dialects (e.g. Northeastern felda “field”, Sinti rada “wheel”, cimera “room”, or lagra “concentration camp” from Feld, Rad, Zimmer, and Lager, respectively). A question arises whether the last type of gender assignment is triggered by a property of German, or whether some specific assignment rules were operating in the Northern dialects at the time of German lexical influence.

The phenomenon of the ‘replacive’ gender assignment needs further investigation, but there are at least a few convincing cases. The genderless Hungarian világ “world” and virág “flower”, for example, have been adapted as the masculine világó and the feminine virágga, respectively, in a variety of Romungro, replacing the older masculine svito “world” and the feminine luludi “flower”, which are attested in closely related Romungro varieties (cf. Elšík et al. 1999).

5.2 Adaptation of loan-words

The pattern developed in Stage B requires that the base form of any integrated loan noun must formally fit in one of the athematic inflection classes. Contrary to what may be expected the adaptation mechanism is quite complex, especially if viewed from the surface perspective of deletion and addition of some phonological material. The clue to the adaptation mechanism consists of taking into account internal morphological structure of the given noun in both Romani and the source language. It is always the stem of the source noun, i.e. not the base form, which is adopted. This becomes apparent if the base form and the non-base-form stem differ (Figure 11).
An interesting phenomenon exists in some Romungro borrowings from Hungarian. If a Hungarian noun has an irregular possessive stem, then it is this irregular stem that is adopted into Romani: e.g. *vő* (possessive stem *ve*-j-) “son-in-law” is borrowed as *vejo*. If, on the contrary, the possessive stem is regular in Hungarian, then the Romani form is based on a non-possessive stem: e.g. *felhő* (possessive stem *felhő*-j) “cloud” is borrowed as *felhő-v-o*.

<table>
<thead>
<tr>
<th>Source language</th>
<th>Base form</th>
<th>Stem</th>
<th>Romani</th>
<th>Base form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serbian</td>
<td>orao “eagle”</td>
<td>orl-</td>
<td>Bugurdži</td>
<td>orl-os</td>
</tr>
<tr>
<td>Albanian</td>
<td>ahër “stable”</td>
<td>ahr-</td>
<td>Bugurdži</td>
<td>ahr-i</td>
</tr>
<tr>
<td>Serbian/Croatian</td>
<td>klinac “nail”</td>
<td>klin-</td>
<td>Romungro</td>
<td>klin-o</td>
</tr>
<tr>
<td>Hungarian</td>
<td>mocsök “bago”</td>
<td>mocsk-</td>
<td>Romungro</td>
<td>moč-k</td>
</tr>
<tr>
<td>Hungarian</td>
<td>majom “monkey”</td>
<td>majm-</td>
<td>Romungro</td>
<td>majm-o</td>
</tr>
</tbody>
</table>

Figure 11: Stem adaptation of loan nouns in Romani

It is significant that the boundary of the inflectional stem in Romani and in the source language almost always coincide. This means that bilingual speakers have access to the internal structure of the source language forms.

5.3 Borrowing new segments and processes

Beside a number of derivational devices Romani also borrows inflectional formants, both segments and processes. A morphophonological process has been borrowed, for example, from Hungarian by a variety of Romungro. In the former language an underlying *yod* triggers gemination of a preceding palatal (e.g. *bátya* “uncle” + 3SG.POSS -ja > *bátyja* “her/his uncle”), and gemination plus palatalization of a preceding dental (e.g. *barát* “friend” + 3SG.POSS -ja > *barájtja* “her/his friend”). Both processes have recently been copied in the Farkašda variety of Romungro: e.g. *rokoñ-i* “relative” + NOM.PL -ja > *rokoñña* “relatives”, or *somsíd-o* “neighbor” + NOM.PL -ja > *somsídída* “neighbors” (but cf. *rokoñja* and *somsídja* in closely related varieties).

In Stage B nominative forms of both numbers were borrowed from the contact language, while forms of other cases were derived by Romani inflectional rules. An important innovation in Stage C is the fact that usually only formants of the plural number are borrowed, while singular, i.e. base form, formants of post-Greek contact languages are either lost through adaptation, or ‘drawn’ into the existing pattern of base form formants (see 4.2). As a consequence of the adapting (rather than borrowing) strategy as far as the base forms
are concerned, the base form formants cannot contribute to an increase of inflection classification. What is surprising is that not even the post-Greek plural formants, which are borrowed (rather than adapted), contribute to such an increase. That is to say, if a post-Greek NOM.PL formant is borrowed into Romani, it almost always replaces an older one. Thus the only effect of the formant replacement is rendering a more salient expression for the NOM.PL. Boretzky & Igl (1991:32) have observed that bisyllabic or long NOM.PL formants are often generalized in Romani, even if they are distributionally marked in the source language – perhaps they may be said to be borrowed selectively.

The ‘saliency’ of the NOM.PL is also provided from internal resources. There are instances of split phonological development of originally identical forms, the NOM.PL acquiring the more salient result. For example, the older Roman form barja, which served as the NOM.PL as well as the ACC.SG of bar “garden”, has developed into barča in the former function, but lost the yod (i.e. bara) in the accusative. Similarly, in Farkašda Romun gro the form kafid’a (of kafidi “table”) has been retained in the ACC.SG, but strengthened into kafid’da in the NOM.PL. The ‘saliency’ of the NOM.PL form may be explained in terms of Bybee’s (1985) relevance principle, which requires greater form differentiation across semantically more relevant categories (number, in our case).

The main source of classification increase in Stage C is the interaction of thematic and athematic segments: cf. the Bugurdži form lovđžida “hunters” consisting of the -d- of the older form lovđžides, and of the thematic suffix -a.

5.4 Intra-paradigmatic levelling

An important part of Stage C developments is exhibited by intra-paradigmatic levelling. For example, in Vlax – and independently also in a few other dialects, e.g. in Harvato – the OBL.PL vowel is assimilated to the singular one in some classes (e.g. OBL.PL *for-en → for-o-n in analogy to the OBL.SG for-o-s, and OBL.PL *romn-j-en → romn-j-an in analogy to the OBL.SG/NOM.PL romn-j-a). This development strengthens the class identity. Similarly, different stem formatives such as jotation may acquire various types of distribution through intra-paradigmatic levelling. An extreme case of intra-paradigmatic levelling is reflected in the athematic masculine classes in Bugurdži:
The intra-paradigmatic levelling has been so extensive (e.g. OBL.PL *daj-en → daj-o-j-en in analogy to NOM.PL daj-o-j-a and SG daj-o-s) that it moved the inflectional stem boundaries. This, in effect, resulted in a classification decrease, since the original classification vowels now form a part of the stem rather than the formant. The athematic o-masculines and u-masculines, for example, now belong to one class. Classification of athematic masculines is now carried out only by the plural suffixes (-j- vs. -d-), and the cumulative NOM.PL suffixes (-a vs. -es).

6. **Stage D**

6.1 Nouns

In some dialects (e.g. in Welsh Romani, Sinti, and the Southern Central Romani) a high degree of inter-paradigmatic levelling has taken place. Levelling between two or more classes may ultimately lead to the extinction of one of them, thus contributing to a decrease of classification. Levelling between thematic and athematic classes may ultimately lead to the decline of the thematicity dichotomy. Let us consider some developments of athematic noun classes.

First, there may be interaction within athematic classes. In Austrian Lovari as well as Slovak-Czech Bougešti (author’s fieldwork) the original athematic i-masculines have taken over the oblique formants of the athematic o-masculines (e.g. doktor-i “doctor”: OBL.SG doktor-os-, i.e. not the original *doktor-is-, and OBL.PL doktor-on-, i.e. not *doktor-en-). In Piedmontese Sinti the class of athematic i-masculines has become extinct, and even the nouns derived by the suffix -ar- – which are the core members of this class in most dialects – have shifted to the o-masculines (e.g. virtaro “inn-keeper”).

---

<table>
<thead>
<tr>
<th>#Mo</th>
<th>*Mu</th>
<th>*MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>“uncle”</td>
<td>“grandfather”</td>
<td>“drill”</td>
</tr>
<tr>
<td>“garden”</td>
<td>“hunter”</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12: Athematic masculine noun classes in Bugurdži
Second, and more importantly, there is interaction between thematic and athematic classes. In the Southern Central dialects, including Roman, the thematic and athematic o-masculines differ only in their OBL.SG and NOM.PL formant. The distinction in the NOM.SG has been lost due to the final slessness and the innovative, predominantly penultimate, stress pattern. The recent development in Roman is towards the loss of the thematicity dichotomy. Animacy of a noun is becoming more important than its thematicity or athematicity. Consider the forms in Figure 13. The nouns in the first and the second columns are thematic, and the nouns in the third and the fourth columns are athematic. In Figure 14 the specifically athe
tematic elements are represented by “A”, the specifically thematic elements by “T”, and elements indifferent as to thematicity by “I”:

<table>
<thead>
<tr>
<th></th>
<th>T.AN</th>
<th>T.INAN</th>
<th>A.AN</th>
<th>A.INAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM.SG</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>ACC.SG</td>
<td>T</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBL.SG</td>
<td>T</td>
<td>T/A</td>
<td>A/T</td>
<td>A (T)</td>
</tr>
<tr>
<td>NOM.PL</td>
<td>T</td>
<td>T (A)</td>
<td>A/T</td>
<td>A/T</td>
</tr>
</tbody>
</table>

Figure 14: Distribution of thematic/athematic elements in Roman o-masculines

It may be observed that the athematic formants infiltrate the thematic paradigms more efficiently with inanimate nouns, and, vice versa, the thematic formants enter more easily into the animate athematic paradigms. Ultimately, if such a development continued, the distinction between the two paradigms would be one of animacy rather than thematicity.

The situation in Welsh Romani is in principle similar to Roman. Here, however, it is only the OBL.SG thematic formant that has expanded to athematic masculine classes (e.g. OBL.SG urě-es- “hedgehog” like čav-es-“boy”, but still NOM.SG urě-os vs. čav-ô, and NOM.PL urě-i vs. čav-ô). An
extreme result of inter-paradigmatic levelling has come about in German Sinti, where the o-masculines of European origin decline exactly like those of Asian origin. This means that the class of o-masculines is unmarked as to the origin of its members (e.g. NOM.SG čav-o “boy” like Sint-o “Sinto, Gypsy”, NOM.PL čav-e like Sint-e, and OBL.SG čav-es- like Sint-es-). Piedmontese Sinti, on the other hand, has retained the thematicity dichotomy, keeping the two classes distinct (cf. NOM.SG čav’-o vs. S’int-o, and NOM.PL čav’-e vs. S’int-i). In the base form, where the inflectional suffixes are identical (i.e. -o), the distinction is carried out by stress position. One may observe in this connection that Piedmontese Sinti has not lost the conservative stress pattern, while German Sinti as decribed by Finck (1903) has.

6.2 Adjectives

A comparison of the athematic adjective inflection with the inflection of the declinable thematic adjectives in Early Romani is given in Figure 15:

<table>
<thead>
<tr>
<th></th>
<th>NOM</th>
<th>OBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG.M</td>
<td>-o</td>
<td>-e</td>
</tr>
<tr>
<td>SG.F</td>
<td>-i</td>
<td>-a</td>
</tr>
<tr>
<td>PL</td>
<td>-e</td>
<td>-a</td>
</tr>
</tbody>
</table>

**Figure 15: Early Romani adjective inflection**

There were two principal lines of development of the Early Romani athematic inflection. A number of dialects (the Northeastern, the Northern Central, the Northern Vlax dialects, and a few Balkan dialects, e.g. Bugurdži, Prilep, and Erli) have retained the specific inflection class. In other dialects, on the other hand, the inflection of European loan adjectives has become identical to the inflection of the Asian o-adjectives (e.g. in the Welsh, Sinto-Manuš, and most Southern Central dialects). In a few of these latter, however, new athematic classes were created after the athematic inflection had merged with the thematic one (e.g. the Welsh Romani blūa “blue” class with SG -a vs. PL -i, or the Piedmontese Sinti tamblo “dark” class with M.SG -o vs. F.SG/PL -i).

Moreover, loan adjectives may also remain indeclinable. The indeclinability may concern all loan adjectives (as in Ajia Varvara, or Sepeči), just a few of them (as in Welsh Romani or Sinti), or a certain stratum of loans. The last case is documented from Roman. Here adjectives borrowed from German, i.e. from the current contact language, do not decline for agreement.
categories. The previous borrowings from South Slavic and Hungarian, on the other hand, inflect like the thematic o-adjectives in Roman. Although the existence of the thematic class of zero-adjectives could play a role, the indeclinability of loan adjectives – rather than being a result of inter-paradigmatic levelling – is a manifestation of the general analytic drift in Romani.

In some dialects athematic adjectives variantly neutralize the category of case, or the category of number. In both instances it is the unmarked forms that take over the marked functions: the nominative forms in Lotfitko and Prilep (i.e. SG -o, PL -a), and the singular forms in Serbian Kalderas (i.e. NOM -o, OBL -on-e). In Hameln and Piedmontese Sinti there was a general loss of case inflection in all adjectives. The process is documented, and it is interesting to see that the loss followed an animacy scale. In German Sinti at the time of Finck (1903), case neutralization was affecting just the agreement with inanimate nouns, while adjectives modifying animate nouns still had the original inflection.

7. Conclusion

The number of Proto-Romani noun classes, some of which have inherited a potential of being lost through ‘natural’ developments (see 3.1), have been considerably increased through contact with Greek at the Early Romani stage. A new adaptation pattern was developed at that time. It gave rise to a strict thematicity dichotomy – a cross-linguistically unique phenomenon (see 4.1), and restricted the possibilities of inflection borrowing in later stages (see 5.3). The marking of classification, which developed as a consequence of the inflection borrowing into Early Romani (see 4.1), not only broke the structure of inherited Layer I segments (see 4.2), but also became the basis for creation of new noun classes (see 4.2). Intra-paradigmatic levelling developments have either strengthened class identity, or lead to classification decrease (see 5.4). The latter was a consequence of inter-paradigmatic levelling, too.

A special type of inter-paradigmatic levelling, the thematicity dichotomy blurring (see 6), is a feature peculiar to some Romani dialects only. The reasons for the decline of athematic morphology remain a puzzle. There seems to be a structural explanation at least for blurring between the thematic and the athematic o-masculines. The loss of the conservative stress pattern plus the final s-lessness make them indistinguishable in their base forms. This means that they constitute complementary classes, and that in one of these two paradigms one more reference form in the lexicon is needed. There is a mutual impact between the thematic and the athematic classes in Roman, but the
development in German Sinti requires that the athematic classes be structurally marked. In Welsh Romani, thematic elements expand even without the existence of complementary classes (since the athematic o-masculines are not s-less, and the stress patterns clearly differ). Thus, the developments in Sinti and Welsh Romani as well as a greater susceptibility of the athematic classes to typological change require a structural or functional account of their markedness, which could be one of the future points of investigation.

Acknowledgements
I wish to thank Peter Bakker, Norbert Boretzky, Jan Červenka, and Yaron Matras for useful comments on the manuscript. Inspiring remarks and questions were provided by the audience of the Fourth International Conference on Romani Linguistics (Manchester, September 1998) where the first version of this paper was read.

Notes
1 Descriptions of the following dialects have been surveyed: Welsh Romani (Sampson 1926), German Sinti (Finck 1903, Holzinger 1993), Piedmontese Sinti (Franzese 1985), Lotfičo (Mánuš et al. 1997), Xaladytko (Wentzel 1980); East Slovak Romani (Hübschmannová et al. 1991), Romungro (Ešák et al. 1999), Roman (Halwachs 1998); Sepeči (Cech & Heinschink 1997), Bugurđi (Boretzky 1993); Ajia Varvara (Igl 1996), Serbian Kalderaš (Boretzky 1994), Taikon Kalderaš (Gjerdman & Ljungberg 1963), Austrian Lovari (Cech & Heinschink 1998), Hungarian Lovari (Hutterer & Mészáros 1967), and Polish Lovari (Pobožniak 1964); and the general description of the dialects of Yugoslavia (Boretzky & Igl 1994).
2 It seems that if there is a split between the OSA and the BSA within a language, then the OSA strategy is preferred by nominals which are higher on an animacy scale. Cf. the oblique suffix -n- in Evenk (Konstantinova 1964) oblique cases of personal pronouns, but not nouns (e.g. si-n- “you.SG” vs. oron- “hart”), or the oblique suffix -walu- in Djapu (Morphy 1983) oblique cases of pronouns and human nouns, but not non-human nouns (e.g. dhuway-walu- “husband” vs. bumbaru- “rock”). This observation needs, of course, a large sample confirmation.
3 I will deal only with the de-jotation of the zero-feminines in this paper, although similar argumentation could be adduced in order to explain the extinction in some dialects of the class of zero-masculines with the morphological number neutralization (MØ-O).
4 Three distinct masculine forms exist in Punjabi and Lahnda (NOM.SG vs. NOM.PL/OBL.SG vs. OBL.PL) as well as in Marathi and Konkani (NOM.SG vs. NOM.PL vs. OBL).
5 Only one feminine form (i.e. a subparadigm with number and case neutralization) exists in Gujarati, Rajasthani, Hindi, or Parya. Punjabi and Lahnda have two forms distinguishing number, and Marathi and Konkani three forms, with the same categorial distinctions as in the masculine subparadigm (see note 4).
6 Prevailing number of Asian noun loans retained at least in some Romani dialects belong to the zero-classes, both masculine and feminine, but there are also instances of i-feminines, o-masculines, and one instance of an i-masculine (cf. Boretzky & Igl 1994:329-332).
In many languages the integration strategy closely adheres to the classical structuralist view of markedness: the marked classes contain only borrowed items, while the unmarked ones are indifferent as to the origin of their members. Unlike these languages Romani has developed a dichotomy in which some classes (the athematic) are marked for non-originality of their members, while other classes (the thematic) are marked for originality. Moreover, there are generally no unmarked classes as far as the origin of their members is concerned.

In some dialects (e.g. in Lovari) the noun papis “grandfather”, the only widespread representative of the α-masculines, became an o-masculine. Less likely it was already borrowed in the form papos (from a different Greek dialect).


In some dialects (e.g. in Southern Central Romani) the thematicity dichotomy blurring also affects such domains as derivation of adjectives, diminutives, abstract nouns, and inchoatives. Sometimes the blurring is bidirectional, i.e. not only Asian grammatical segments may be used with European lexical items, but also European segments may be used with Asian items. In deverbal abstract noun derivation only the first direction of blurring occurs.

Halwachs (p.c.) informs me that the accusative form grof-o, which is not mentioned in Halwachs (1998), also exists as a rare variant in Roman.

Although weaker proficiency may have played a role in the developments in Welsh Romani and Roman, it cannot generally explain the phenomenon of the dichotomy blurring and the decline of athematic morphology. For instance, there are thousands of proficient speakers of the Southern Central dialects today, and surely there were centuries ago, when a number of manifestations of the thematicity decline in these dialects must have come into being.

References


