Eötvös Loránd Tudományegyetem
Bölcsészettudományi Kar

DOKTORI DISSZERTÁCIÓ

Baló András Márton

Analógia a lovári alaktanban
(Analogy in Lovari Morphology)

Nyelvtudományi Doktori Iskola
Dr. Tolcsvai Nagy Gábor akadémikus, egyetemi tanár

Elméleti Nyelvészeti Doktori Program
Dr. Bánréti Zoltán Csc.

A bizottság tagjai és tudományos fokozatuk:

A bizottság elnöke:
Dr. Siptár Péter DSc.

Hivatalosan felkért bírálók:
Dr. Törkenczy Miklós DSc.
Dr. Zsigri Gyula PhD.

A bizottság tagjai:
Dr. Ladányi Mária DSc.
Dr. Szalai Andrea PhD.
Dr. Cser András PhD.

Témavezető:
Dr. Kálmán László CSc.
Budapest, 2016

ADATLAP
a doktori értekezés nyilvánosságra hozatalához

I. A doktori értekezés adatai

A szerző neve: Baló András Márton
MTMT-azonosító: 10038065
A doktori értekezés címe és alcíme: Analogy in Lovari Morphology
DOI-azonosító: 10.15476/ELTE.2016.171
A doktori iskola neve: Nyelvtudományi Doktori Iskola
A doktori iskolán belüli doktori program neve: Elméleti Nyelvészeti Doktori Program
A témavezető neve és tudományos fokozata: Dr. Kálmán László CSc.
A témavezető munkahelye: ELTE BTK, MTA Nyelvtudományi Intézet

II. Nyilatkozatok

1. A doktori értekezés szerzőjeként

a) hozzájárulok, hogy a doktori fokozat megszerzését követően a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az ELTE Digitális Intézményi Tudástárban. Felhatalmazom az ELTE BTK Doktori és Tudományszervezési Hivatal ügyintézőjét, Manhercz Mónikát, hogy az értekezést és a téziseket feltöltse az ELTE Digitális Intézményi Tudástárba, és ennek során kitöltse a feltöltéshez szükséges nyilatkozatokat.

b) kérem, hogy a mellékelt kérelemben részletezett szabadalmi, illetőleg oltalmi bejelentés közzétételéig a doktori értekezést ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

c) kérem, hogy a nemzetbiztonsági okból minősített adatot tartalmazó doktori értekezést a minősítés (dátum)-ig tartó időtartama alatt ne bocsássák nyilvánosságra az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban;

d) kérem, hogy a mű kiadására vonatkozó mellékelt kiadó szerződésre tekintettel a doktori értekezést a könyv megjelenéséig ne bocsássák nyilvánosságra az Egyetemi Könyvtárban, és az ELTE Digitális Intézményi Tudástárban csak a könyv bibliográfiai adatait tegyék közzé. Ha a könyv a fokozatszerzést követőn egy évig nem jelenik meg, hozzájárulok, hogy a doktori értekezésem és a tézisek nyilvánosságra kerüljenek az Egyetemi Könyvtárban és az ELTE Digitális Intézményi Tudástárban.

2. A doktori értekezés szerzőjeként kijelentem, hogy

a) az ELTE Digitális Intézményi Tudástárba feltöltendő doktori értekezés és a tézisek saját eredeti, önálló szellemi munkám és legjobb tudomásom szerint nem sértem vele senki szerzői jogait;

b) a doktori értekezés és a tézisek nyomtatott változatai és az elektronikus adathordozón benyújtott tartalmak (szöveg és ábrák) mindenben megegyeznek.

3. A doktori értekezés szerzőjeként hozzájárulok a doktori értekezés és a tézisek szövegének Plágiumkereső adatbázisba helyezéséhez és plágiumellenőrző vizsgálatok lefuttatásához.

Kelt: Budapest, 2016. december 12.

a doktori értekezés szerzőjének aláírása

Contents

Part Two

List of tables

List of figures

Part One

1 Introduction

1.1 Background, aim and methods

In the present dissertation,[1] I will attempt to explain some particular instances of **variation** in the morphology of Lovari, a dialect of Romani belonging to the Vlax dialect group. This will be done in an **analogical framework**, relying only on surface forms and their relationships, using the notions of constructions and **schemata** (Goldberg 1995, Booij 2010); underlying forms of any sort or abstract levels will not be posited.

As language is very complex, I would not like to postulate rules and constraints that are as general as possible; instead, I would like to grasp some minor phenomena and the

1 The dissertation was written as part of the project Variation in Romani Morphology, supported by the Hungarian Scientific Research Fund (OTKA, Project 111961, project leader: László Kálmán). Although I am alone responsible for the content of this dissertation, several people helped me with my work, and I thank them all. I hope I will not leave out anybody, but if I do, it will only be my fault. First and foremost, I would like to thank my supervisor, László Kálmán for launching the OTKA project within the framework of which I could write the dissertation, for reading the drafts and the papers that preceded them, for making a lot of comments and suggestions, and especially for coming up with great ideas; without his help, the content of this work would be much poorer. I would also like to thank the two opponents, Miklós Törkenczy and Gyula Zsigri, for meticulously reading the first version that was handed in, giving me a lot of useful advice based on which I could revise the dissertation, and Miklós Törkenczy also for allowing me to consult him when it was necessary.

My thanks also go to Kinga Gárdai, without whose assistance I would have had a much harder time dealing with the administrative and technical matters; Balázs Surányi and Huba Bartos for their help with the OTKA project proposal; István Kenesei and the Research Institute for Linguistics for the background; Andrea Szalai, Krisztina Polgárdi and Ferenc Kiefer for their support; Péter Rebrus for the discussions we had about related and unrelated topics, as well as his useful insights and the articles he directed me to; Michael Beníšek for the discussions on Romani nouns; András Komlósy for his kind and serious professional attention; Mátyás Arató for his expert participation in the fieldwork; Yaron Matras and Viktor Elšík for involving me in their Romani-related projects; and Zsuzsanna Bodnár for introducing me to the know-how of fieldwork.

Some of the ideas presented in the dissertation were aired at conferences such as the International Conference on Romani Linguistics, the AnaMorphoSys Conference, the International Morphology Meetings, the Décembrettes International Conferences on Morphology and the Poznań Linguistic Meetings. I would like to thank the people there who listened to my presentation and gave me feedback on my work, among others Marcin Kilarski, Alice Harris, Gregory Stump and James Blevins.

possible forces behind them. Although neurolinguistics is still in its infancy, based upon recent research in the field (Menn & Duffield 2014) it seems that construction-based and **usage-based** approaches to grammar can provide insights into how grammars can come closer to reflecting what our brains do. Therefore, I will adopt the view that 'the human mind is an inveterate pattern-seeker' (Blevins & Blevins 2009: 1). Patterns are generalised through analogical reasoning, and cognitive psychological research also justifies the idea that analogical reasoning is an essential part of human thinking (Penn, Holyoak & Povinelli 2008). This "**non-analytical**" approach[2] is also in line with recent experimental research in phonetics, speech perception and speech production (Port 2007, Port 2010). Apparently, in speech perception 'the data strongly suggest that listeners employ **a rich and detailed description of words**' (Port 2007: 145) instead of abstract, segmented forms. In other words, 'listeners encode particulars rather than generalities' (Pisoni 1997: 10).

In morphology, this represents itself in the comparison of **surface forms**, which is a basic tenet of an analogical framework. While the notion of output-output correspondence has been present in Optimality Theory, the theory still applies general principles and constraints (even if they can also be unranked, cf. McCarthy 2008), and they still rely on the theoretical notion of input, even if the inventory is defined by the outputs, and there are no restrictions on the input (this is called the richness of the base, Prince & Smolensky 2004: 205).[3] There have also been attempts at explaining variation (e.g. Boersma & Hayes 2001, Anttila 2002) within the framework of Optimality Theory, but they cannot say much about the real reasons and proportions of variation.[4]

2 About a more holistic view and the presence of the analytic approach in linguistics, see Kálmán (2007).

3 Rebrus & Törkenczy (2005), in order to account for the appearance of a phonologically and/or morphologically unexpected allomorph, specified the input only morpho-syntactically, but not phonologically.

4 A thorough overview of the connection of analogical frameworks to other theories, including the advantages of an analogical approach over generative and other, more traditional approaches, as well as analogical models such as AML (Analogical Modelling of Language, Skousen 1989) or TiMBL (Tilburg Memory-Based Learner, Daelemans & van den Bosch 2005), is provided by Rung (2011).

1.2 Overview of the structure of the dissertation

The dissertation consists of two main parts. The first part (Chapters 1-3) is of an introductory nature, while the second part (Chapters 4-7) presents the research itself.

In Chapter 2, I will provide **an introduction to the Romani language**. This includes its history, its research, as well as an outline of its dialects. These will be discussed both on an international level and with regard to the Hungarian scene. This chapter also contains the description of the current research, fieldwork and material on which the dissertation is based. In Chapter 3, I present the notion and significance of **analogy in grammar** in more detail, taking the following aspects under scrutiny: patterns, rules and categories; the role of similarity and frequency; prototypes and paradigms; and the idea of rich memory.

After a brief overview of the **Lovari sound system** in Chapter 4 at the beginning of Part Two, I will provide my analysis of the **Lovari nominal and verbal system** in Chapter 5, while also presenting the existing analyses and correcting them where necessary. In Chapter 6, I will introduce the nature of **weak points**, present the notion of **schemata** as used by Booij (2010), create an **improved version** and apply them to three different instances of **variation in Lovari morphology**. Finally, in Chapter 7 I will draw the conclusions and go through the possible directions of further research.

2 Introduction to the Romani language

2.1 Discovery, early history and research

Until very recently, no attempts were made at codifying the Romani language, neither in terms of standardisation, nor in the sense of creating a uniform written form. It has always been an **oral language** and it is still one chiefly, which gives it a special status among the more codified and standardised languages of Europe. The Romani people are known in Europe as nomadic travellers, or the lowest stratum of the sedentary society of the country, or both. All descriptions of the *romaní šib* 'Romani language' have come from **external sources**. The first traces of Romani in Europe are to be found in the 16th century – a few sentences and short lists of words. Sources from the 17th century are scarce but exist (for instance, from Evliya Çelebi in his famous travelogue), and it wasn't until the late 18th century that scholars managed to establish the Indo-Aryan origins of the language and its kinship with the languages spoken in India beyond a shadow of a doubt. Hungarians were in the vanguard of research with István Vali (Wáli), son of a Calvinist pastor, who, while studying in the Netherlands in the mid-18th century, recognised the similarity between the language spoken by Ceylonese students and the language he had heard at home spoken by the Gypsies (Landauer 2004). The late 18th century saw several other scholars (Heinrich Grellmann, Johann Rüdiger in Germany, William Marsden, Jacob Bryant in England) becoming engaged in the study of the language of Gypsies, and in 1782, probably partly based on the findings of Vali, Johann Rüdiger published his ground-breaking essay, *Von der Sprache und Herkunft der Zigeuner aus Indien.*

Rüdiger noticed the lexical **similarities between Romani and Hindustani** and with the help of a native speaker of Romani and a description of Hindi, managed to prove that the two languages are also similar morphologically and syntactically (Bakker & Matras 1997). The father of Romani linguistics is August Pott, who published his work, *Die Zigeuner in Europa und Asien*, a concise grammar and dictionary in 1844-45 based on several descriptions of **diverse Romani varieties** spoken in Europe. He established the fact that the different varieties go back to **one root**, and that due to **language contact**, there are different linguistic layers to be found in the language, which gave some clue about the possible migration of Gypsies.

The fact that at some time or another Greece must have been part of the migration route was first stated by Franz Miklosich in his comprehensive work published between 1872 and 1880, *Über die Mundarten und Wanderungen der Zigeuner Europas*. Based on meticulous study of different Romani dialects and comparing them to Indo-Aryan languages he concluded that the Gypsies must have left India after the disappearance of the Sanskrit nominal case system (which will be a crucial part of our discussion), about the 10th century. The **Greek element** was established based on lexical borrowings, and an important aspect of the discovery was that these borrowings are shared by all dialects, so the dialectal dispersion must have been preceded by a lengthier Greek sojourn.

We still do not know exactly why the Romani people left India; whether this took place in several waves; and when exactly. There is written evidence (and not only from the poet Firdusi) of musicians from India, called *luri*, going to Persia, invited officially, but the connection between the *luri* and the Romani people is not proven. However, there is linguistic evidence in the lexicon of Romani presence in Persia and Armenia, too. **Iranian and Armenian** loan-words are shared by all Romani dialects, just like Greek ones, so there is no reason to doubt that the Romani people still constituted one group when they arrived in Europe.

Pott (1844-45) already suggested that the Romani people are related to commercial nomadic castes of India (Matras 2002: 15). The word ḍ*omba*, originally designating a low caste of travelling musicians and dancers, appears to be a cognate of *řom/rom*. The ḍ*omba* are still an existing group in several regions of India and Pakistan (Schmid 2007). The word and the people are not only related to the Romani people, but also to other groups of Indian origin living outside India. The two other notable groups, the Domari and the Lomavren people are also marginalised groups where they live: the Domari are scattered in the Middle East (for the variety spoken in Syria, see Herin 2012, for the variety spoken in Jerusalem see Macalister 1914 and Matras 2012), the Lomavren live in Armenia. These groups are related in several aspects. Their **ethnonyms (*řom, dom* and *lom*) are cognates**, their languages are related Indo-Aryan languages spoken outside India, and they are all called historically Gypsies (*tsoʿanim* for the Dom in Jerusalem, a blend from Yiddish ציגיינער *tsigayner* and Hebrew צוֹעֲן *tsoʕan*, and Armenian Gypsies for the Lom – cf. Finck 1907). Both are endangered, if not already extinct languages. The non-sedentary nature of these groups is also marked by the common term of these languages for people outside the

group (*gāžó* in Lovari), which also has the additional sense ‘peasant’. Ralph Turner and John Sampson conducted a debate in 1926-27 (Turner 1926, 1927 and Sampson 1927) raising the question whether the social similarity of these groups is also a proof of a common origin and a common ancestral language. There is still no conclusive historical evidence, however, for these claims, and we still do not know either, when and why the Dom and the Lom left India.

The Persian lexical elements in Romani[5] (e.g. *zōr* ‘strength’ < *zūr*, *baxt* ‘luck’ < *baxt*, *ambról* < *amrud*, *angrusťi* ‘ring’ < *anguštarī*, *armaja* ‘curse’ < *armān*, *korro* ‘blind’ < *kūr*, *mol* ‘wine’ < *mol*, *čerhaj* ‘star’ < *čarkh*, *res-* ‘arrive’ < *rasīdan*), as well as the historical accounts of musicians or dancers from India arriving in Persia point to the fact that the Romani people spent some time in Persia, while the lack of Arabic elements suggests that they left before the Muslim conquest.

The weight of the Greek influence on the language makes us think that the Roma must have arrived in Byzantium by the 11th century, before they scattered across Europe from the 13th-14th centuries onwards. All this means that they could have left India as early as the 5th century and moved on from Persia around the 7th or 8th century, although some, like Matras (2002) suggest a later outward migration from India, around the 8th-9th centuries (the lack of Arabic influence can also be a result of little contact with the ruling Arabs in Persia), while others, like Kaufman (1984) place it as early as around 300 BC, although this suggestion lacks any solid foundation.[6]

The time in Armenia is also treasured in the lexicon, although not as abundantly as the time in Persia (e.g. *čohají* ‘witch’ < *čivagh*, *grast* ‘horse’ < *grast*), or at least not in the varieties spoken in Hungary. The reason for this is not clear, especially because according to Boretzky (1995a) the Roma must have spent a longer time in Armenia, too. From Armenia, the Romani people arrived in Byzantium, where the main language, the language of trade was Greek, which heavily influenced Romani, not only its lexicon but its grammar, too. Even if the Roma spoke Romani among themselves, they had to learn the language used there, too; just like later in Hungary and all over Europe; monolingual Romani

5 Although it is not necessarily easy to distinguish the Persian borrowings from the appropriate Persian cognates shared by Indo-Iranian languages in general.

6 Matras (1996) suggests that the different lexical components do not necessarily tell us exactly the route of migration, as most of the lexicon not inherited directly from Indo-Aryan could have been picked up in Anatolia, which was mainly Greek-speaking during the Byzantine period, with Persian and Armenian in close proximity.

speakers virtually do not exist, and bilingualism must have started when they left India.

Apart from Indo-Aryan, the biggest part of the lexicon common to all Romani varieties derives from Greek. In addition to that, Romani adopted completely new and very characteristic parts of its grammar from Greek. The influence of Greek on Romani grammars suggests that their stay in Byzantium must have been fairly long and formative. This is also partly speculation, as the earliest document that proves their actual presence comes from the 1280s: a letter concerning tax collection "from the so-called Egyptani and Athingani". Based on the legend about the *luri*, the Romani people have sometimes been associated with the Lori people, another nomadic tribe descending from India. In the course of history, there have been other groups with whom the Roma have been identified. In the Byzantine Empire, one of them was the Athinganoi, a sect which had probably disappeared by the time the Roma arrived in Byzantium. The Athinganoi were mentioned in religious texts for the first time around 800 AD. They are people who tell others' fortunes and try to influence others with their unchristian teachings. The appearance of the Roma, perhaps because they were similarly mysterious or because they carried out similar trades, could have evoked the memory of the Athinganoi, and the name "stuck". We cannot make straightforward assumptions based on exonyms about the people they describe, but in the case of the Athinganoi, it seems likely that the term referred to the Roma, but it probably referred to another group originally. According to some researchers, the commonly used exonym for the Romani people, Hungarian *cigány* and its cognates in other languages (Serbian *цигански*, Slovakian *cigán*, German *Zigeuner*, Italian z*ingaro* etc.) derive from the name of the Athinganoi. The Greek word Ἀθίγγανοι originally means "people who do not want others to touch them, who do not want to be touched, who are untouchable". Hübschmannová (1972) suggests that the origin of the name as used for the Roma might be found in the commands of cleanness, as they are common in India; thus, the Romani people would have distanced themselves from the rest of the population in certain respects or, at least, were found to be different. But these conclusions are not generally accepted. According to Matras (2011), it seems more plausible that the word *cigány* and its cognates are related to medieaval personal name forms Scygan, Zygan, Zegan, Zigan, Chygan, Czygan, Cygan, Cigan, Chigan and place name forms Zygan, Zigan, Cygan, Cyganuaya, Cziganvaya, Chiganvaya, Czyganwaya, Chiganwaya, Chyganfalu, Czyganfalwa, Czynganfalwa, Chyganfalwa, as well as present-day Cigánd in

Hungary and Cigányi (Crişeni) in Romania (Nagy 2004). Fehértói (1987) claims that these early place names and personal names (most probably pronounced [siga:n] and [tʃiga:n]) come from Old Turkic *sïγan* as glossed in one example in the Old Turkic Dictonary: *sïγan sač* 'smooth hair', or from an Old Turkic name for low-caste slaves, *čïγan*.[7]

From Italian and German travellers we know for sure that from the 13th century onwards Romani people lived around the city of Methoni, half-way between Venice and the Holy Land, and thus an important station of pilgrims travelling to Palestine, on the Peloponnese. According to the traveller Arnold von Harff (1471-1505), they lived in poor huts and many of them were skilled smiths (Gilsenbach 1994), but no mention is made of their language. By that time, so the end of the 15th century, the Roma had started to migrate further into Europe. Their unusual appearance, their non-sedentary lifestyle and their own language which they would not give up were not well-received wherever they went. **From a linguistic perspective, this is the point where the diversification of dialects must have begun.**

It is very difficult to reconstruct the spatial, temporal and formal dimensions of the Romani language as a more or less uniform entity, as spoken before we have any attestation of the diverse European dialects (Matras 2002: 18-20). The name used for the hypothetical proto-variety spoken after the differentiation of New Indo-Aryan languages and before the Byzantine period is **Proto-Romani** and it is a sum of the changes that took place before the Roma reached Europe. We have no written records of this period, but hypothetical forms can still be reconstructed if necessary. For example, a feminine demonstrative pronoun **ota* can be reconstructed based on: 1) the equivalent *ola*, which still persists in some dialects, and the renewed form *odola* in other dialects; 2) its correspondence to related forms in related languages, like Domari *ora*; 3) the regular change of Old Indo-Aryan /t/ to /l/ in Romani and /r/ in Domari.

Early Romani is a term used for the stage the language reached after the Greek influence but before the spread of dialects. It is also hypothetical to a great extent, but inferences are easier to make here. Besides a large amount of loanwords (e.g. *drom* 'road' < *δρομος*, *eftá* 'seven' < *εφτα*, *fó̋ro* 'town' < *φορο*, *xōjí* 'anger' < *χολη*, *karfin* 'nail' < *καρφι*, *lulud^{j}í* 'flower' < *λουλουδι*, *mesají* 'table' < *μεσαλι*, *skamín* 'chair' < *σκαμνι*), heavy

7 The other common term for the Roma, *Gypsy* and its cognates (French *gitan*, Spanish *gitano*) comes from the Greek word *Αἰγύπτιοι* because once it was mistakenly believed that they came from Egypt (Fraser 1992).

morphological borrowing took place from Greek into Romani. This has often happened to Romani due to the circumstances in which it is used, so Romani is a rich field for those who are interested in the effects of language contact. **Borrowing mostly happens from the local language**, that is, the one spoken by the majority society surrounding the community of the Roma, and has been particularly intensive since the appearance of the Roma in Byzantium and their dispersal in Europe. A phonological aspect of Early Romani might have been the phoneme /ř/ in *řom* 'man', which could have been the uvular /ʀ/, still preserved in some dialects, merged with /r/ in others, or the Proto-Romani retroflex /ḍ/ (Matras 2002: 20).

2.2 Later history

The first written records of the presence of the Roma in Europe outside the Greek territories, in Transyilvania and Paris come from the early 15th century (Davies 1996). **The dispersal of the Romani in Europe also marked the diversification of dialects**, due to internal development and language contact. According to the sources, the 16th century already saw their cruel persecution in Germany, Italy, Sweden, England, Denmark and other European countries (Fraser 1992, Kenrick 2007). Migrant and itinerants are never truly welcome, and the Roma were especially conspicuous because of their different external traits, especially in Western Europe. The letter of safe conduct allegedly issued by Sigismund, Holy Roman Emperor in 1417 appears to be a fake (Nagy 2004). Their reception was slightly different in the Balkans, where people were more used to migrant groups, and in Russia, where they were left in peace provided they paid their taxes; but they were kept as slaves for centuries in the historical regions of Wallachia, Transylvania and Moldavia. According to some researchers, almost two-thirds of the Romani people of Europe currently live in the territory of South-Eastern Europe (Dupcsik 2009), with significant Romani population in Bulgaria, Hungary and Romania. The Romani people are also present overseas; their immigration began in the early 19th century with the Romanichals from the United Kingdom. Following the abolition of slavery in Romania, the late 19th century saw the arrival of Eastern European groups both in North and South America, and there are Romani communities even in Australia (cf. Fraser 1992; Salo &

Salo 1986), including groups speaking Lovari, where they arrived via Western-Europe as a result of a second wave of migration from Eastern Europe (Romania, Serbia, Hungary, Slovakia). This continued in the 1990s, after the fall of communism.

The situation of minorities in Western Europe has always been more difficult to tackle because of the different approach to ethnic and national minorities, as opposed to Eastern Europe, where, at least in theory, minority affairs are part of the government agenda. The integration of the Romani people has been a hot topic in Hungary and several other countries in the region, and the solution seems to be far too distant. The lack of long-term thinking on a state level, based on a complex view of historical, sociological and demographical aspects is a serious problem. The forced settlement of the Roma ordered by Maria Theresa in 1758 confronted them with the need to change their lifestyle completely and abruptly, but such a change can only be achieved through smooth transition. While an itinerant way of life is something that Europe still cannot handle after centuries, many Romani groups could not adapt to a sedentary lifestyle, perhaps due to its forced nature. The result is very unappealing, according to my personal experience gained whilst doing fieldwork in Hungary and based on personal discussions. Although there are several attempts to keep, maintain and improve the Romani culture, tradition and language in Hungary, as the general attitude of the majority of the society towards the Roma is hostile or neutral at best, the Romani people themselves do not think their language is of great or, even worse, of any value. The slightly more ambitious members of the community find that assimilation is the only way to break out of the trap of poverty and prejudice, while the marginalised majority of the Roma, the outcast in the countryside are still stuck at the age when they were forced to settle down and they are still at their wits' end as to how to move on. Stripped of their roots, they have no means which would help them with their integration and a possible transition from their previous itinerant state to a settled one.

The language of this group, **the Romani language is still a living language in many parts of Europe** but probably not for very long. Language shift is happening at an enormous speed, with the majority of the younger generations not being able to speak at all (see the details under the discussion of Romani dialects). But, in spite of all the hardship, the extensive borrowing and language contact, the core of its structure still survives in its present-day dialects.

2.3 Para-Romani

The very first source on Romani comes from an English traveller and physician, Andrew Boorde, collected in 1542 and published in 1547 in his *The Fyrst Boke of the Introduction of Knowledge* (Boorde *et al.* 1870), 'a handbook of Europe, Barbary, Egypt and Judæa'. He presents 13 sentences of the 'Egipt speche' (Boorde *et al.* 1870: 217-218) along with their 'Englyshe' translation. The sentences were probably collected from Romani people living in England (Eliav-Feldon 2012: 123). One example can be seen in (1).

(1) I wyl go wyth you. *A vauatosa*. (Boorde *et al.* 1870: 218)

Although the word boundaries were not established quite adequately, this is very much like what is spoken today. The same sentence in present-day Lovari is given in (2).

(2) *avav* *tusa*
'go' 1ST SING. PRES. IND. 2ND SING. PERS. PRON. INSTR.
'I will go with you.'

Although the typical scenario for Romani since the morphological borrowing from Greek is that the structure is preserved but there is extensive lexical borrowing. In some cases, however, like in the case of Romani spoken in the United Kingdom, things went the other way: after the disappearance of Romani as an every-day language in England and Wales in the late 19th-early 20th centuries due to **language shift**, some Romani words are still kept and used in the framework of English grammar, as it is illustrated in (3).

(3) *mandi* *has delled* *a*
1ST SING. PERS. PRON. LOC. (!) 'give' 3RD SING. PRES. PERF. ART. INDEF.
mush
'man' NOM. SING.
'I have hit a man.' (Matras 2010: 121)

This blend is called *Romanes* ‘Romani’ by its speakers and **Para-Romani** by the literature (Matras 2010, Bakker & van der Voort 1991, Bakker 1998 etc.). **Both terms are somewhat misleading**, however. The first one because it is not Romani but English interspersed with vocabulary of Romani origins, used as a kind of special lexicon; the second one because it has nothing to do with paralanguage as formulated first by Trager (1958). Besides Anglo-Romani, there are other similar varieties, like the Caló languages spoken in Spain and Portugal (Ackerley 1914 and 1929, Bakker 1995, Ignasi-Xavier 2005, McLane 1985, Boretzky 1992) Scando-Romani (Johansson 1977, Hancock 1992, Ladefoged 1998) but Lomavren, mentioned in section 2.1, is also similar in that its grammar is Armenian whereas its lexicon is almost entirely of Indo-Aryan origins. According to Kovalcsik & Kubínyi (2000), a similar variety called Hungaro-Romani exists in Hungary as well, spoken by the Roma who went through a language shift generations ago.

The status of Para-Romani has been discussed extensively (Matras 1998 and 2010, Hancock 1970 and 1984, Boretzky 1985 and 1998, Courthiade 1991, Thomason & Kaufman 1988), but it is still not clear whether to consider it a creole, a mixed language or something else. Tálos (2001), for instance, maintains the creole theory. For Para-Romani languages, however, **the basic grammatical structure is always formed by the dominant language**, and the vocabulary comes from Romani. Creoles, on the other hand, are generally defined (Holm 2000) as languages with a lexicon from the dominant language (the “base language”) but the grammatical structure originating from the other language. With this definition in mind, **Para-Romani varieties are definitely not creoles**. We should also add that, according to recent research, creoles are structurally distinct from non-creole languages, as Bakker *et al.* (2011) and McWhorter (2005) claim, but t**he structure of Anglo-Romani is very similar to that of English**, which is considered a “proper” language.

According to the literature, the level of intertwining in the case of mixed languages is so high that it is not possible to define one single ancestor language (Matras & Bakker 2003). In the case of Para-Romani varieties, this is not the case, so they are best considered mere styles of speech. Based on Hancock (1984), we can add that their existence may have been triggered and maintained because it is **a means of reinforcing an in-group identity**.

2.4 Recent research

Recent research into Romani has focussed mainly on the following aspects:

1) The unity and diversity of Romani dialects and their implications for both dialect classification and linguistic origins;
2) the impact of language contact on linguistic change, including grammatical borrowing and contact-induced internal innovation, as well as the retention of Romani vocabulary in instances of language shift;
3) the sociology of the language, in particular questions of status, codification, and standardisation;
4) Romani in the context of current theoretical issues in general linguistics. (Bakker & Matras 1997: ix)

As for domain 1), **the unity and diversity of Romani dialects**, the obscurity of Proto-Romani and Early Romani, as well as the question how uniform they were make historical Romani linguistics a very difficult field.[8] Comparative Romani dialectology, on the other hand can provide us with the opportunity of widening **the scope of variation** and looking into the aspects and meaning of variation not only within a dialect but also across a language with many distinct dialects. Domain 2), **the impact of language contact on language change**, is also a difficult topic as there is often too much emphasis placed on the opposition between internal or inherited and external or borrowed in Romani linguistics. Domain 3), **the sociology of the language** is not a linguistically relevant question in the strict sense; nevertheless, it raises important questions and provides interesting answers. Finally, we must note that in spite of its presence on the list, the last domain, **Romani in the context of theoretical linguistics**, is sadly lagging behind compared to the other three, with not so many, although very valuable papers (Friedman 1991a, b, Plank 1995, van der Auwera 1998 with a more traditional approach, and more recently Elšík & Matras 2006 and Baló 2012, 2015).

8 Although see a well-grounded reconstruction of Early Romani in Elšík & Matras (2006).

2.5 Dialects

Within Romani, some 1000 lexical roots form the Early Romani inheritance, out of which about **700 hundred are of Indo-Aryan origin**. However, all 700 hundred roots rarely appear simultaneously in one single dialect (Matras 2002: 21). According to Vekerdi (1971a), up to 80 roots are missing from Hungarian Lovari, and this number, although we have no precise statistical evidence, must have risen since then. Nevertheless, this component of the lexicon is still the source of basic vocabulary. As mentioned in section 2.1, there is a shared set of items from **Persian and Armenian, all in all about 100**. In addition to the Indo-Aryan core, some **200-250 Greek roots** are also shared by all dialects, but again, this number does not actually mean that all of these roots are present and actively used in all dialects; it is much lower for individual dialects, and the roots can rather be described as an overlapping set. Although some (e.g. Boretzky 1992) exclude the Greek elements from the core lexicon, as they are common to all Romani dialects, there is no real reason for not regarding them as its basic part. **Morphological borrowing** also took place from Greek, but this is quite natural if we consider for how long the Roma stayed in Byzantium. Although exposed to the effects of language contact, the morphology of Romani is also quite uniform across dialects. The split into dialects took place fairly late in history, and, as the British Romani example in (3) shows, the basics are still preserved. Its phonology and syntax are often more prone to change under the influence of the surrounding language.

Being **a spoken language spread over a large territory**, Romani is also **variegated**; there are certainly grammatical differences among the different dialects, and, as the language has never been officially and cross-dialectally reformed, the **extensive lexical borrowing** has made each dialect one of its kind. The Romani groups living in a certain country usually borrow lexical items from the language used by the surrounding community or country.

There have been attempts at **translation on an international level** but this poses several problems. How shall we create the words that do not exist in the language? They are frequently supplemented from English or Latin, but will that necessarily make it more understandable for all Romani speakers? Which dialect should we use? On the one hand, the dialects described in this section can be mutually intelligible, but at the same time there

can be so many differences between two, fairly distinct dialects that **the speakers might switch to a language that they both speak alongside Romani** (Boretzky 1995b). Constant migration, albeit somewhat surprisingly, can also be a cementing force, as the Romani people meet again and reunite, and often **the lingua franca is Romani**, despite the various effects of the different majority languages.

There have been various attempts to classify the dialects of the Romani language. Based on structural differences, the currently most accepted framework divides it into **dialects as branches of proto-varieties** (Miklosich 1872-80, Bakker & Matras 1997), which emerged after the Byzantine period, when the Romani people began to disperse in Europe.

> [The viewpoint] is further reinforced by an assumption that, since Romani lacks coherent and continuous territorial representation, its dialects are not subjected to a geographical diffusion pattern of innovations, and consequently do not form a geographical dialect continuum. (Matras 2005a: 7).

The proto-varieties established based on the branching model are split into **further varieties due to further migration**; thus, for instance, whereas Lovari was originally spoken in western Romania, it is possible to talk about Hungarian and Austrian Lovari, which coexist and interfere with the Romungro and the Burgenland Romani varieties, which are spoken in the same area, respectively, but belong to a different dialect group.[9] **The branches correspond to geographical locations only roughly**, partly due to the lack of the study of cross-dialectal variation in Romani, so Matras 2005a challenges the branching model and approaches the current situation from a geographical point of view (see this section further below).[10]

9 To what extent are the dialects and dialect groups mutually intelligible thus depends on both linguistic and social factors. Although the two dialects I have the most experience with, Lovari and Romungro belong to different dialect groups, their speakers can understand each other, while finding each other's dialect somewhat odd; the question is rather whether they are willing to understand each other.

10 There are other, less plausible and less justified suggestions for the classification of Romani dialects, like Kaufman (1979) or Tálos (2001). Kaufman (1979) partly corresponds to the branching model, adding geographical aspects, too, but mistakenly placing genetically related and geographically close dialects under distinct branches, with as many as seven branches deriving from a common ancestor called "Proto-European Romani" and including the Americas, too. Tálos (2001) classifies the dialects into three main groups, but his classification, although intended to cover them on an international, or at least European

Based on the reference grid of the branching model of Romani, the language can be divided into **four main groups called Balkan, Vlax, Central, and Northern**. The Northern branch is less coherent than the others, used more like an umbrella term for the varieties, spoken not only in the north, but also in the west and south of Europe, as well as some other varieties considered extinct by now. The standard, branching dialect classification can be seen in Table 1, based on Bakker & Matras (1997), Matras (2002), the referenced literature therein, as well as Boretzky (2003) for the Vlax dialects and the Linguistic Atlas of Central Romani (http://ling.ff.cuni.cz/atlas/) for Central Romani.

The Balkan dialects are sometimes subdivided into northern and southern Balkan. The names of subgroups and individual dialects are often derived from Turkish. Muslim Roma often refer to themselves as Xoraxane, meaning "Turkish Gypsies", from *xoraxaj* 'foreigner'. These dialects are characterised by a probably longer-lasting and therefore stronger Greek influence and a heavy Turkish influence as well.

The general view of Vlax dialects is that they are the most widely spoken, both geographically and numerically, and the best documented. Although there are many publications, including textbooks and descriptions about Vlax internationally (cf. e.g. Hancock 1995, Lee 2005), **the lack of actual data, both with respect to quality and quantity, both internationally and otherwise, is striking**. Even Boretzky (2003), a comprehensive work on Vlax Romani, can only draw on a small amount of data which are either old or consist of fairy tales or contain pre-written texts (Boretzky 2003: 3-11). In the case of Hungarian Lovari, the sources are collections and descriptions by József Vekerdi and a few others (Vekerdi 1961, 1966 and 1985, Hajdu 1960, Mészáros 1968, Valis 1968), and, most unusually, a Romani textbook (Choli-Daróczi & Feyér 1988). Even Matras (2002) claims that 'we have fairly extensive documentation' (Matras 2002: 8) on Lovari, but only cites sources such as Ackerley (1932) (!), Pobożniak (1964) and Cech & Heinschink (1999), all of which are old and/or mere descriptions or collections of words. Thus, we must say that, contrary to general belief, **not even the Vlax dialects are so well documented**.

---

level, is based on the three dialects spoken in Hungary. This leads to a rather arbitrary grouping of fairly different dialects. For example, the variety spoken in the Ukraine is considered a Vlax dialect by Kaufman (1979) and a Central dialect by Tálos (2001), whereas the dialect spoken in Hungary and called Romungro is classified as the relative of Balkan dialects by Kaufman (1979) but not so closely related to the Carpathian dialect spoken in Slovakia, although they are both belong to the Central dialect group, as rightly pointed out by Tálos (2001).

	Balkan	Vlax	Central	Northern
where it is spoken	Turkey, Greece, Bulgaria, Macedonia, Albania, Serbia (Kosovo), Romania, Ukraine and Iran	Serbia, Montenegro, Croatia, Bosnia-Herzegovina, Macedonia, Southern Romania, Bulgaria, Greece, Albania, Turkey, Romania, Moldova, Hungary	Hungary, Slovakia, Northern Slovenia, Eastern Austria, Ukraine, Czech Republic, Poland, Russia	Germany, Austria, France, Italy, Netherlands, Belgium, Finland, Poland, Lithuania, Latvia, Estonia, Russia, Belarus, Ukraine, Hungary, Czech Republic, Slovakia, Russia
individual dialects	Arli, Erli, Mečkar, Sepeči, Ursari, Crimean Romani, Zargari, Romano, Drindari, Kalajdži, Bugurdži	Agia Varvara, Gurbet, Džambazi, Kalburdžu, Čergar, Kalderaš, Lovari, Čurari, Mačvaja	Romungro, Vend, Burgenland Roman, East Slovak Romani, Bergitka, Prekmurje, Bohemian Romani, West Slovak Romani	Sinti-Manuš, Finnish Romani (Kaale), Laiuse Romani, Xaladitka, Polska Roma, Čuxny (Loftiko), Abruzzian, Calabrian, Dolenjski, Istrian Romani

Table 1

The dialects of Romani

Some things can nevertheless be established about Vlax dialects in general, for example that **they share a lot of Romanian lexical and even morphological elements**. It is indeed widely used in Europe, having spread after the abolition of slavery in Romania in the second half of the 19th century. Vlax is also divided into a northern and a southern group, with the southern dialects spoken in Greece and the Balkans, while the northern dialects, like Lovari and Kalderaš, spoken in Romania, Hungary and other neighbouring

countries. The Lovari and Kalderaš dialects can also be found in Western Europe and the Americas.

The Central dialects are also divided into two, northern and southern branches. While the northern Central dialects have been spoken in the territory of the former Czechoslovakia, the southern Central dialects, Romungro and Vend were once widely spoken in Hungary (Vekerdi 1981, 1984 and, more recently, Bodnárová 2013). Although the Romungro and the Lovari varieties belong to different branches, it is not at all impossible for speakers in Hungary to understand each other's speech, if there is a will.

The term Northern, used for the fourth group, is more of a collective umbrella term for dialects spoken mainly in the west and north of Europe, but also in the south, from Wales to Finland and Italy. Para-Romani varieties, discussed in section 2.3, show traces of dialects that once belonged to this group (British Romani and Iberian Romani), but the most central of this dialect group are the Sinti-Manuš varieties spoken in Germany and France, respectively. It is believed that the diverse varieties appearing all over Europe spread from Germany. According to some sources, like Mészáros (1980), Sinti was once spoken by a small group of people in Hungary as well. Along with Finnish Romani, these dialects belong to the Northwestern group, while the Northeastern group comprises those spoken in Russia, Poland and the Baltics. Abruzzian Romani and the variety spoken in Slovenia are usually categorised under the Northern heading, although the former seems to be more closely related to the Balkan dialects (Matras 2002), while the latter shows the features of several dialect groups.

We must also mention two varieties spoken in Hungary, the Cerhari and the Gurvari (Mészáros 1976 and Vekerdi 1971b) dialects, which are generally considered transitional varieties of the Central and the Vlax dialect groups.

Some genetic features set apart the Vlax dialects from non-Vlax dialects indeed, like the genitive ending: *-k-* in Vlax and *-ker-* in non-Vlax (for example Central) dialects. Similar, well-known features of Vlax dialects as opposed to non-Vlax dialects are the reduction of affricates to sibilants (*šāvó* 'boy' in Vlax and *čhāvó* in Central, Northern and Balkan dialects; also *ža-* 'go' versus *dža-*) and the first person singular marker in the past: Vlax *-em* as opposed to Central *-om* (*phendem* 'I said' in Vlax versus *phendʲom* in Romungro). Lexical features can be mentioned, too: *kor* 'neck' in Vlax as opposed to *men* in the other dialects (Bakker & Matras 1997).

Complementing and partly replacing the branching model which emphasises genetic criteria, Matras (2005a) sets out **a geographical diffusion model** and outlines several isoglosses existing Europe.

> According to this model, relations between dialects are not absolute, based on 'genetic' criteria, but relative: dialects are more closely, or more remotely, related to other dialects, depending on the number of relevant features that they share. The structural features that distinguish dialects are a result of processes of change and innovation that spread from one community to another. The outcome of these changes can be plotted on a map in the form of isoglosses. Dialects thus form a geographical continuum which reflects the historical spread of structural innovations (as well as the clustering of archaisms) in time and space. (Matras 2005a: 8)

The genetic approach relies on observations regarding shared features and their temporal aspects, but the same **features can also be viewed on a spatial basis**. However, the historical approach so essential to the genetic point of view plays a role in the location of diffusion spaces as well, as the geographical isoglosses established in this manner are said to have emerged between the 15th and 17th centuries, following the emigration from the Balkans but probably before another wave of migration and dialect displacement.

Matras (2005a) further argues that, as Romani linguistic structures are distributed geographically, and the first noteworthy sources about the language and its different dialects come from the 18th century, the diffusion and differentiation of linguistic developments took place during a period of settlement in the 16th-17th centuries. In this framework, **features which are shared by certain dialects but not shared by others are not results of earlier ties but of their geographical proximity**. The classification of dialects should not be based on the common features retained from Early Romani, but features that actually emerged as innovations within a geographically definable network of speakers.

As Romani is a most diversified language and is constantly diversifying, it is no surprise that people try to find a systematic way in order to grasp this diversification. In this aspect, it seems to be a good idea to separate common features from shared features, and their geographical distribution can obviously help. An example for **a common feature**

is the raising of the vowel in the conjunction *te* > *ti*, which is found in Sinti spoken in Germany and Austria as well as Sepeči spoken in Turkey, so two, geographically distant dialects. **A shared feature**, for instance, in a geographically coherent area is the retention of past forms of verbs inflected for gender, like *gelo/geli* 'he/she went' (e.g. in the Kalderaš dialect, Boretzky 1994: 71) on the one hand, and the lack of gender marking *gēlas* 'he/she went' (e.g. in the Lovari dialect, author's own collection). The two geographically coherent areas are separated by an isogloss, which is part of a bundle of isoglosses called the Great Divide (Matras 2005a), and crosses Central Europe between the lines Southeastern Austria-Hungary-Romania in the north and Slovenia-Croatia-Vojvodina and the Danube in the south. This means that **the genetically related Vlax dialect group is divided geographically**, while two, genetically distant dialects, like Sinti and Lovari may share the same feature.

Some features have been considered to be genetic, supporting the existence of the branches summarised in Table 1, like the alternation of /s/ and /h/ in the copula (e.g. *me som* 'I am' in Vlax as opposed to *me hom* in Sinti), in an intervocalic position in grammatical markers (*lesa* 'with him' in Vlax and *leha* in Sinti) (cf. Matras 2005a: 18-19) and word-finally. However, it seems that /h/ instead of /s/ in the copula and in an intervocalic position occurs in a coherent geographical zone ranging from France and including Germany, Northern Italy, Central Europe and Finland. According to Matras (2005a), the intervocalic /h/ always coincides with the /h/ in the copula. He also adds that the the /s/ in a final position was lost in a zone along the eastern coast of the Adriatic Sea, spreading to such dialects as Doljenski, Arli, the Southern Central group and southwestern Vlax. As for the Southern Central group, this is definitely true, based on the author's own fieldwork; one of its trademark features is the replacement of /s/ with /h/ intervocalically and the complete loss of /s/ word-finally. However, the data I collected calls into question the validity of both the genetic and the geographical model. **It seems that neither genetic, nor geographical aspects play a role** in the variation of word-final /s/ and /h/. The word-final sound of the 3rd person singular past indicative verb forms in (4) alternates among /s/, /h/ and /x/ without any apparent reason.

(4) *prāsalás* *les* *addigra*
‘mock’ 3RD PERS. SING. PAST IND. 3RD PERS. SING. PERS. PRON. ACC. ‘until then’ ADV. ‘
mīg *opré* *či*
when’ ADV. ‘up’ PREP./VERBAL PART. ‘no’ ADV.
xojajdáx *les*
‘GET/BE ANGRY’ 3RD PERS. SING. PAST IND. 3RD PERS. SING. PERS. PRON. ACC.
taj *tēlé* *pusadáh*
‘and’ ADV. ‘down’ PREP./VERBAL PART. ‘stab’ 3RD PERS. SING. PAST IND.
les
3RD PERS. SING. PERS. PRON. ACC.
‘He kept mocking him until he got angry and stabbed him.’

To conclude this section, we have to say that **both the genetic and the geographical model are appealing** as models and can serve as a reference grid. They are also similar in many ways, trying to point out structural features based on which Romani can be classified into dialects.

2.6 The Romani people in Hungary

The most thorough research to date concerning the history of the Romani people in Hungary has been carried out by Pál Nagy. According to his research (summarised primarily in Nagy 1998, 2004), the very first source that mentions the presence of some Romani people in the territory of Hungary is found in some Brașov (present-day Romania) accounts from 1416, where they are referred to as Egyptians. The first charter that actually talks about Roma (*ciganos*) was issued by John Hunyadi, then perpetual count of Beszterce/Bistrița to Péter and Tamás Barcsay and allowed them to keep four Romani people as serfs on any of their estates freely (Nagy 1998).

There are more and more data concerning Romani people in the territory of Hungary from the 1470s, but their number must have been still low at the time, and they concentrated mainly in the larger towns of Transylvania, Brașov, Cluj and Sibiu (Nagy 2004). In a charter issued by Matthias Corvinus in 1487 the Roma are allowed to keep their

own customs. **In these sources, however, no references are found regarding their language.** Instances of the presence of Romani people outside Transylvania are first mentioned in the late 15th and early 16th centuries, but no reference is made to significant, settled groups. According to one of the sources, Romani musicians played in front of Queen Beatrice, wife of Matthias Corvinus in 1489.

The arrival of the Romani people in Hungary was not organised or large-scale, and it took place in waves, influenced by the political and historical processes of Europe. A higher number of people migrated to Hungary in the 16th and 17th centuries, when their prosecution in Western-Europe began (see section 2.2 and Nagy 2004). This had a major and a minor consequence: the major one was that the Roma arriving from the south did not continue their journey to Western Europe; the minor one was an eastward migration to Hungary. The term *germani zingari* is first used in the statute of Sopron county in 1717 (Nagy 2004), and it probably referred to the group we call Sinti.

As their migration into Hungary was gradual and came from several directions[11], and due to the fact that it is not mentioned in the sources, **it is almost impossible to tell which variety of Romani they spoke**. Nagy (2004) claims that Romani was not uniform by the time a larger number of Romani people arrived in Hungary. However, if the geographical model outlined in section 2.5 (Matras 2005a) is closer to how dialect diversification actually took place, then all Roma spoke a more or less uniform language at the time. This is corroborated by Nagy (2004) (somewhat contradicting himself), who also claims that the different Romani groups (and varieties of the Romani language) known today did not exist as such between the end of the 14th and the end of the 17th centuries in Hungary, and no homogeneous professional groups existed either.

The most likely scenario is that the group commonly referred to as Carpathian or Romungro, who have spoken a Central Romani dialect, arrived over a longer period, between the 15th and 17th centuries, whereas the migration of the Vlax group and the group known as the Boyash happened between the 18th and 20th centuries (Nagy 2004, Achim 2004). If the differentiation of dialects also took place between the 15th centuries and 17th centuries, as the geographical model suggests, then **the speech of the Romungro was not so different on their arrival from that of those later called as Vlax**; it only became

11 The migration from the south came mainly from Serbia and Bulgaria to the southern counties (many of them accompanying the Turks), and partly from the same places, partly from Transylvania to the central counties. Their appearance in the northern and northwestern parts of Hungary came even later.

different later in an areal manner (note that most of the territory of present-day Slovakia, where the Northern Central Romani variety is spoken, was part of the Kingdom of Hungary at the time).

The demand of the society that the Roma should settle down was not as straightforward between the 15th and 17th centuries. But even at the time, there were those who led an itinerant life, but there were also those who had already led a settled life. There was a transitional group, too, whose members became part of the social and economical structure of a given village, but did not give the seasonal summer migration (Nagy 1998).

The term *zingari valachi* 'Vlax Romani' appears first in the early 18th century in Hungarian sources. By this time, their dialect must have been different from the dialect of those who had already been living in Hungary. Their arrival enhanced the hostile feelings of the local population towards the Roma, while, at the same time, the Hapsburg monarchs made attempts at their forced assimilation. In a royal edict, Joseph II **forbade them to use their mother tongue**, reinforcing a previous edict issued by Maria Theresa. Whoever spoke Romani was to receive a punishment 24 strokes by caning.[12] In the 19th century, most of the Romani people of Hungary were leading a settled life; they were those who were considered "normal", as opposed to those, mostly Vlax Roma, who still led a mainly itinerant life. The lifestyle of the settled ones were not much different from their non-Romani counterparts. According to the 1893 census, only 30 per cent of the Romani of population of Hungary at the time said their mother tongue was Romani. Although the territory was different, the proportion is telling (Dupcsik 2009).

2.7 Romani dialects in Hungary

The diverse Romani groups have not changed much since then, except that even the small number of itinerant Roma have settled down, too. The most recent demographic survey on was carried out in 2003 by István Kemény and his colleagues on a representative sample (c.f. Kemény & Janky 2003). They estimate the number of Romani people living in Hungary at 520,000-650,000. This means that **the Romani population has been steadily**

12 In addition to that, of course, they were forbidden and compelled to do several other things. They were not allowed to migrate, they could not wear their traditional clothes, they had to go to church and their intermarriage was also encouraged by financial means.

growing, from 320,000 in 1971 and 468,000 in 1993.[13] The three groups they took into consideration were the Romungro, the Vlax Romani and the Boyash groups (whose language is a Romanian dialect, not a Romani one). In their survey, they also asked questions about the mother tongue. Considering the present-day territory of Hungary in light of the 1893 census mentioned in section 2.6, the proportion of the speakers of Romani was only 10 per cent, as opposed to Transylvania, where it was 42 per cent. The 1971 survey carried out by Kemény and his colleagues showed a significant increase: Romani as a mother tongue was present in 21.2 per cent of the population. However, by the 1993 survey this number dropped dramatically to 4.4 per cent (Kemény 2000). This increased slightly in the ensuing ten years and reached 7.7 per cent. This suggests that after the assimilating tendencies of the socialist era in Hungary, we can see some kind of return, or at least a desire to return to the roots.

We must note that **the linguistic aspects of these demographic surveys are somewhat sketchy** and confuse the categories of group belonging and mother tongue, as Szalai (2007) also points out correctly. Kemény & Janky (2003) state that the Romungro group includes those who only speak Hungarian, whereas those who are bilingual in Romani and Hungarian belong to the Vlax group, and those who are bilingual in Boyash and Hungarian belong to the Boyash[14] group. But while the categories of group belonging and mother tongue often correspond to each other, generally they only overlap. Therefore, if we can take the percentage seriously, that 7.7 per cent of the total Romani population of Hungary whose mother tongue is Romani must contain all the Romani varieties spoken in Hungary. Theoretically, **these varieties include the Romungro and Vend dialects** belonging to the Central dialect group (referred to as Carpathian Romani in some of the relevant literature), **the sundry Northern Vlax varieties**, for example Lovari, as well as the **Sinti dialect** (Northern dialect group) and the transitional varieties like **Cerhari and Gurvari**.

Although most of the people belonging to the Romungro group[15] went through **a**

13 According to the most conservative estimates, they number about 3.5 million in the whole of Europe and a further 500,000 in the other parts of the world (Matras 2005b).

14 Material on the language of the Boyash, an archaic Romanian dialect is scarce; there is no international literature to note. Regarding the Boyash language as spoken in Hungary see Orsós & Kálmán (2009) and Arató (2015).

15 This is the group traditionally linked to musicianship; they usually play the typical and traditional "Zigenuermusik". On the other hand, traditional Romani folk music nowadays is rather played by Vlax Romani people.

language shift already, there are places in Hungary where the dialect is still spoken (cf. e.g. the Linguistic Atlas of Central Romani and the author's own fieldwork, as well as earlier sources like Vekerdi 1981 and 1984): several villages in Nógrád country, some towns in the area of the capital and many villages and towns in the south and west of the Transdanubia region.

On the other hand, **those who call themselves Vlax do not necessarily speak the language**, as it is revealed again by the fieldwork I conducted. From a linguistic aspect, only that 7.7 per cent of the Romani population of Hungary, who speak Romani, is of interest.[16]

While the Romungro, Vend and Sinti groups are considered more or less uniform, the Vlax Romani people of Hungary are said to constitute a more diverse group.[17] Somewhat contradicting the relevant state-of-the-art international literature, different, self-designated Vlax tribes are often identified with different linguistic groups, and every tribe is said to correspond to one single dialect (Erdős 1959, Tálos 2001): Lovara, Colara, Kalderaša, Cerhara, Mašara, Fodozovo, Romano rom, Bodoca, Kherara, Bugara, Čurara, Patrinara, Drizara (Erdős 1959: 33).[18] As we can see, the list is very mixed, with some varieties already mentioned as dialects that exist in some way (Lovari and Kalderaš, two Vlax dialects, as well as the transitional Cerhari and Čurari varieties), while the rest are only known from Erdős (1959), Tálos (2001) and occasional self-designation.

Boretzky (2003), for example, mentions the fact that the Romani textbook Choli-Daróczi & Feyér (1988) contains inflectional forms that are not typical of Vlax: the dominant first person singular present indicative form of the copula is given as *som* instead of *sim* (which is indeed typical of Lovari, as confirmed by the author's fieldwork), and the dominant first person perfective marker as *-om* instead of the typical *-em* (also confirmed by the author's fieldwork). Boretzky (2003), based on personal communication with József Vekerdi, suggests that this is due to Choli-Daróczi's mixed, Lovari and Mašari parentage, and the forms *som* and *-om* come from Mašari (Boretzky 2003: 4). This argumentation is

16 It is also a question whether we can confirm what Szalai (2007) suggests, who still mentions Sinti among the dialects spoken in Hungary. With only a few hundred speakers when it was last described in Mészáros (1980), it can easily happen that it has disappeared since, especially considering the heavy language shift that took place between 1971 and 1993 in the country.

17 The groups who speak the transitional dialects Gurvari and Cerhari are also considered Vlax, although it is more likely that their varieties were not Vlax originally. Tálos (2001) also suggests that Gurvari is an umbrella term for transitional dialects like Cerhari and Čurari.

18 The tribal names are often based on the jobs the groups did.

rather contradictory, however. If Mašari is a Vlax dialect, then it should not matter whether one's parents are Lovara or Mašara, and it should be included in a comprehensive description of the Vlax dialects, such as Boretzky (2003). But if the above forms are not typical of Vlax, and rather derive from Mašari, then Mašari cannot be a Vlax variety, contrary to what Erdős (1959) claims. This is an important instance of of contradiction with regard to my research as well: whether *-om* and *-em* are variants of the same thing, and when the different forms are observed, it is morphological variation; or is it a different, larger group of people who say *som* and another, larger group uses *sim*?[19] Recent fieldwork shows that although minor differences might exist among the self-designated Vlax groups listed above, like the voicing of word-initial /k/ in demonstrative pronouns in Drizari, so Lovari *kado* ~ Drizari *gado* 'this', these differences are probably not enough to consider them separate dialects. Elsewhere, Boretzky (2003) mentions that the Roma in Norway are from a mixed background, too, namely Lovari and Čurari, but then adds, regarding the latter, that 'the entire character as well as the Hungarisms present in the dialect suggest that this is essentially Lovari' (Boretzky 2003: 5).

Generally it is said that the most widely spread Vlax dialect in Hungary is Lovari (from Hungarian *ló* 'horse', derived from its "v-adding" stem alternant, *lov-*, with the addition of an agentive suffix). Endre Tálos in his introduction to *Zhanes romanes?*, the first Romani textbook in Hungarian (Tálos 1988) notes that Lovari is considered a standard variety among the speakers of Vlax Romani, and even speakers of other Vlax varieties adopt and start using it instead of their vernacular (but see the discussion above concerning the other Vlax varieties). Therefore, and perhaps also because it is one of the most common Northern Vlax dialects on an international level, the publications (textbooks, language exam materials etc.) that followed, as well as the official state language exam use Lovari. The identification of the Romani language with its dialect called Lovari, which is a common misunderstanding in Hungary, is not problematic because of the varieties spoken by other Vlax Romani groups (the dialect status of which is questionable), but because of the fact that the other major dialect groups (with regard to Hungary, especially the Southern Central dialects) are ignored.

19 Variation may also be influenced by the geographical location of two groups, both of whom nevertheless refer to themselves by the same autonym. In the case of a language like Romani, where there are so many varieties which are seemingly related and different at the same time, variation should perhaps be considered true variation only on the level of a single speaker.

2.8 Dialect diversity and dialectal pluralism

There are a few factors that influence **dialect diversity and mutual intelligibility** in Romani (cf. among others, Matras 2005b). I will also comment on them and state why these factors do not pose a real problem in any way.

1. Romani is primarily a spoken language. There is no literary version or written standard to which speakers can resort to. Recent Romani literature is an artificial creation, a forced action taken to save face among the other European languages that possess a long written tradition and, through that, elaborate literature, having evolved over hundreds of years at least. As Tálos (2001) notes, every language with a literary was a "barbarian" language once, and the development of a literary standard was triggered by social processes. However, neither the value, nor the use of a language has got very much to do with its literature. The existence or lack of literature does not make a language more or less valued in the eyes of outsiders, and does not make it more viable, either (cf. Ancient Greek with abundant literature, now transformed and extinct, or English, whose literature is probably not richer than that of other European languages).

2. By now, all Romani speakers are bilingual and frequently integrate words from the language of the majority society.[20] This can cause problems in international communication. But this has always been the case with Romani, and it has survived up to the present day as a fully-fledged language. The fact that when meeting Romani speakers from other countries, one must resort to the "core" elements, can actually make the language more stable through a strong sense of group belonging. In addition, many loanwords (see section 2.1) have already become part of the "core" and exist in all varieties.

3. Contrary to what one might think, Romani has always been used to communicate within one's own close group: an extended family or a smaller community. Northern Vlax

20 Szalai (2007, 2015) refers to the bilingualism of Romani people as extended diglossia, based on Ferguson (1959), Fishman (1967) as well as Halwachs (1993). As suggested by the diminishing number of speakers, there is a real threat of language shift among the less educated speakers, who see their language inferior and an obstacle of social mobility. However, Romani intellectuals and non-Romani researchers of the Romani language and culture make every effort to re-introduce the language and emphasise its values to the descendants of those who once spoke it. This can lead to such odd situations where group identity becomes fuzzy: for example, a member of a community which originally spoke a Central Romani dialect will learn Lovari, a Vlax Romani dialect at school.

varieties, like Lovari and Kalderaš can be extremely useful in that respect, as these groups are territorially the most wide-spread in Europe and overseas. This is why, as mentioned in section 2.7, Lovari has come to be some kind of a standard in Hungary as well, and this is why these varieties are probably used for international communication more easily. This kind of migration can pose problems at first, when meeting the communities already living there and speaking a different variety, but the need to communicate can help bridge the linguistic gap. According to Matras (2005b), 'Romani intellectuals especially acquire the skill to handle conversations without resorting to insertions from their respective second languages, and patterns of mutual accommodation in the choice of words and even grammatical structures can be observed' (Matras 2005b: 5).

4. There is no spoken standard; regional dialects prevail. This raises the question of the need for a standard variety, a linguistic norm. There have been attempts to standardise Romani, in order to achieve both uniformity and unity, both linguistically and nationally or ethnically, but none of them have been successful. Gilliat-Smith (1960) thought "Basic Romani", a common, standard variety was theoretically possible, to which came the answer in Wolf (1960), who claimed there was no means or motivation for such a thing (cf. also Matras 1999). But the group of activists and linguists centring around the International Romani Union have never abandoned the idea. Some of them promote their own dialect as a standard (Kochanowski 1995, Hancock 1993), while others suggest that a common European Romani language already exists, it only needs to be rectified and codified, and all dialects can provide parts of this common Romani language (Courthiade 1989).

These well-meaning efforts, however, have several flaws. From a practical aspect, Romani dialects are much less influenced by an adjustment process that can be seen in case of other languages, especially because their speakers are not in contact with each other (Bakker & Rooker 2001). From a theoretical aspect, standardisation and language planning have had strong ties to nationalism (Wright 2004), but a classic, territorial nationalism has no basis for Romani. Therefore, **the standardisation of Romani 'differs from conventional standardisation** in that it shows an effort to forge a shared identity by relying on the symbolic support of transnational organisations of governance and on embracing pluralism of forms' (Matras 2015: 297).

When there are no strong, centralised institutions for the creation, development and promotion of a standard national language, activism is important. The **non-centralised,**

bottom-up approach and the transnational nature of the process, which is an obvious consequence of the lack of any territorial unity, encourages **diversity within uniformity**, that is, achieving the acceptance of the existence of a Romani identity by connecting the Romani communities of Europe and by emphasising their diversity (cf. Matras 2015).

Diversity requires a pluralistic approach, offering space to regional varieties as well as codification efforts, and ultimately the increase of domains where Romani is used. Publications in Romani have been appearing in Europe since the 1970s, and the use of Romani has finally reached the domain of online communication as well (Leggio 2013, Halwachs 2012). **A new form of linguistic variation appears in the virtual domain**: when the Romani language website of the Swedish state radio (Radio Romano) presents Romani news reports from all over the world, they do not attempt to unify the orthography or the language of the audio recordings and so readers and listeners encounter several dialectal varieties (Matras 2015).

> Online social media represent a new domain of language use, one that relies on basic literacy skills and so on a bottom-up codification of language without either a regulatory norm or any form of territorialisation. They show how pluralism of form can exist side-by-side with community-specific dialect choices. (Matras 2015: 303)

Comments posted on YouTube videos (Leggio & Matras 2013) show again a most diverse mixture of local dialects and spellings, while strengthening the existence of a transnational Romani language and a transnational, yet informal Romani identity. This organic process is complemented by more official, politically oriented efforts to gain recognition for the language in Europe. The Council of Europe has been an advocate of the cause since the appearance of written Romani, recommending to give the Romani language an equal status, just like other, regional minority languages have in as early as 1983 (cf. Bakker 2001). This has been followed by reinforcements of the recommendation in the same vein, like in 2000, when it was stated that the opportunity to learn in the mother tongue should be guaranteed for speakers of the Romani language as well. Thus, **the Romani minority has quickly become a linguistic minority in Europe**, and as such, the teaching of the language has also become a priority. Matras (2005b) recommended support for the pluralistic approach, and the Council of Europe went on to state that standardisation is not essential to the maintenance and promotion of Romani.

In Hungary, although pluralism is apparently present on the level of the language as spoken by the small number of native speakers, there is a tendency to see Lovari as some form of a standard or normative variety, due to the reasons discussed in section 2.7.

2.9 Current research activities

The activity of the Council of Europe also encourages the development of teaching materials, and, in order to do that, research and education related to Romani. From the 1970s onwards, a renewed interest is shown in several aspects of Romani linguistics on an international level. One of the places where the most intensive work has been done recently is the University of Manchester. Here, all fields from descriptive linguistics through dialectology, sociolinguistics, language planning and language policy to an online dictionary of Romani dialects and Romani in online communication have received a lot of attention. Besides linguistic topics, they also focus on questions like language and identity, migration and language teaching and offer an extensive online database of almost all aspects (http://romani.humanities.manchester.ac.uk/). Another hub is the University of Graz, where the focus is on how the Romani language, history and culture can be integrated an educational framework and curriculum, aimed at teachers and others occupational groups that deal with the Roma. Their fact sheets (http://romafacts.uni-graz.at/) created with the support of the Council of Europe also provide useful information for those whose intention is to make non-Romani people more aware of and more sensitive to Romani culture, as public knowledge of them is still poor, in spite of the fact that they constitute the largest minority in Europe. According to their credo, education of both sides is the key to integration.

Despite all the goodwill and even professional expertise, the two most relevant projects, the **Romani Morpho-Syntax Database** (Manchester, http://romani.humanities.manchester.ac.uk/rms/) and the **Romani Lexicon Project** (Graz, http://romani.uni-graz.at/romlex/) each have their drawbacks. The projects, according to their description, aim to provide a comparative description of Romani dialects on the one hand, and offer almost complete coverage of the basic lexicon of the Romani language and contain data that are representative of the variation in the lexicon of all Romani dialects on

the other. The projects and their online outlets are supposed to provide a whole-scale, comprehensive view of Romani dialects.

While the projects tried to adopt an all-encompassing approach to Romani, which is a praiseworthy and remarkable effort, the negatives consequences are obvious as well. Firstly, both projects drew on existing published sources, whose reliability and to what extent they are up-to-date are always questions that should not be ignored. Secondly, the additional fieldwork carried out as part of the projects, in which the author participated to a small extent as well, was not as extensive as it could have been and it was limited in both quality and quantity. One dialect or areal variety was covered with one interview in the RMS Database, and the custom word-lists used in the ROMLEX project were not designed to assess the actual situation of the language; some parts of it were aimed to elicit made-up words and expressions.

2.10 Research of Romani in Hungary

As for Hungary, Vekerdi (1982) gives a thorough account of Romani related research in Hungary up to that point. Romani word lists and even grammatical descriptions were written as early as the beginning of the 19th century. The following list is not intended to be exhaustive, I will only highlight a few milestones.

The thorough grammar and dictionary of János Szmodics from 1827 describes the Southern Central dialect of Zala county. János Bornemisza published an analysis of the Southern Central dialect of Nógrád county in 1853. Ferenc Sztojka's dictionary from 1886, commissioned by Archduke Joseph, is the first example of a dictionary which abounds in words artificially created by the author whose mother tongue was a Vlax Romani dialect, interspersed with misunderstandings in the form of mirror translations from Hungarian, like *prahicko žuvlʲi* 'peasant woman' (Vekerdi 1982: 3) from *práho* 'dust' based on two similarly sounding Hungarian words, *por* 'dust' and *pór* 'peasant' (arch.), and actually meaning 'dust woman'. Vekerdi (1982) says that the life oeuvre of Henrik Wlislocki (on the travelling Roma of Transylvania), an often quoted researcher of all aspects of Romani culture is not necessarily reliable.

The comprehensive grammar of Archduke Joseph from 1888 proved to be another

milestone and a reference point for later studies. A huge gap followed until the appearance of Kamill Erdős in the 1950s, a self-taught researcher who tried to classify the different Romani, especially Vlax Romani varieties spoken in Hungary. Finally, it was József Vekerdi and György Mészáros who, based on their own fieldwork, did extensive work in the field of the description of the different Romani dialects existing in Hungary from the late 1960s to the 1980s, with Vekerdi's later addition of a comprehensive dictionary of the Romani dialects in the 1990s (Vekerdi 1971b, 1981, 1984 and 2000, Hutterer & Mészáros 1967, Mészáros 1969-70, 1976 and 1980, Vekerdi & Mészáros 1980).[21]

All in all, **linguistic studies of Romani in Hungary confined themselves to dialectological research and descriptive works** up to 1990. The fall of the communism brought an end to this and opened up the scene for other areas like **theoretical linguistics and sociolinguistics**. It also brought about the emergence of self-proclaimed experts of Romani whose work helped little to improve and increase the knowledge about Romani, only confused the real language as spoken by native speakers with well-meaning lexical and grammatical creations which are, unfortunately, often pure inventions.

2.11 The current research

As can be seen from this chapter so far, trying to do theoretically oriented linguistic research into Romani is not an easy task. **Authentic and trustworthy corpora as such, of any variety of Romani, have not existed until very recently.** If we look at the international landscape, the situation is better now, with small corpora of Thrace Romani-Turkish-Greek and Finnish Romani-Finnish focussing on issues of language contact. Nevertheless, Adamou (2016), adopting a corpus-based approach to language contact while making use of these corpora among others, also mentions that 'one of the main problems for a corpus-based study of lesser-known and endangered languages is the small size of the corpora' (Adamou 2016: 14).

21 From the 1970s to the 1990s, Zita Réger did sociolinguistically oriented research in Romani communities (cf. e.g. Réger 1974, 1990 and 2002).

2.11.1 Background

The bulk of available Romani language material in Hungary is mostly old, outdated and not much, or not authentic enough to be reliable (e.g. Bari 1990). Transcribed versions of collected tales (e.g. Vekerdi 1985) have appeared, but the genre of story-telling can be so different from every-day, casual speech that it is not necessarily wise to use them if we want real life data. If we want to focus on just one dialect, the situation is even worse. Although Lovari is said to be well-documented, this is limited to dictionaries and grammars, only a few of which, referred to in section 2.10 already, are reliable. Out of the 62 tales published in Vekerdi (1985), only 18 are in the Lovari dialect, and 2 are mixtures of Lovari and another variety. Besides the newly collected material, I used these tales for data, as they are, despite everything, still the most reliable sources.

I looked out for material on the international scene as well, and it turned out that the sources available were not of significant amount. Lovari material of Hungarian relevance appears in Cech *et al.* (1999) and Fennesz-Juhasz & Heinschink (1999), but they did their fieldwork in Austria. Although some of the informants, or rather their predecessors came from Hungary, they have lived in Austria at least since the Hungarian Revolution of 1956, but more often since the 19th or early 20th century. Although I scrutinised Cech *et al.* (1999) and Fennesz-Juhasz & Heinschink (1999) carefully, during the research I only turned to them occasionally to double-check. I was given access to the original transcriptions of the recordings which served as a basis for the latter two publications,[22] and they supported the claims I make in the second part of the dissertation. The Romani language material collected by Zita Réger is being processed,[23] but it is expected that no significant amount of every-day, adult speech will be found on the tapes.

The **collection of new, authentic, up-to-date, real life data** began in 2015 within the framework of the project Variation in Romani Morphology, supported by the Hungarian Scientific Research Fund (OTKA, Project 111961, project leader: László Kálmán).[24] Through this data collection we have access to reliable data, but **the process of data collection is made difficult by several factors**, so the amount of data available at the

22 I am extremely grateful to Barbara Schrammel-Leber and the Plurilingualism Research Unit of the University of Graz for their kind assistance.

23 The material gained renewed attention thanks to Zoltán Bánréti and Andrea Szalai, partly at the author's initiative.

24 The fieldwork was carried out by Mátyás Arató and the author.

moment is still not big.

1. **The lack of reliable starting points.** The last geographically oriented data concerning Romani dialects in Hungary comes from Mészáros (1969-70).

> This language geographical description is of a representative nature. It provides an outline of the villages and towns of Hungary where the mother tongue of Gypsies is either Romani or Hungarian or Romanian, while it is also suitable for observing the “linguistic boom” that took place in the Romani language. (Mészáros 1969-70: 309)

Nothing is said, however, concerning the ways and methods of the survey and data collection. The map offered in Mészáros (1969-70) was one of the points of our departure, but several locations have proved to be blind alleys, perhaps because the speakers have disappeared since the survey or perhaps because the original data were not reliable in the first place.

2. **The availability of native speakers.** This refers to a more complex set of problems. Firstly, it is almost impossible to find native speakers in a big city or town without prior knowledge of their whereabouts. That is why the main target of fieldwork is the countryside generally. Secondly, due to the community service introduced recently in Hungary, people in the countryside can be found at home less and, once found, they are less willing to assist after a long day’s work. Thirdly, according to our personal observation, Vlax Romani people tend to be less accessible and approachable than Boyash or Central Romani speakers. This might be a consequence of the different treatment they received over the course of history, as touched upon in sections 2.2 and 2.6: itinerant Roma have always been looked upon with more suspicion than settled Roma, and this could have led to their being more suspicious of strangers, too. Another consequence is that they are less helpful and, in addition, more financially oriented on a number of occasions.

Thus, besides the use of the dialect map in Mészáros 1969-70, one possibility is to go on fieldwork accompanied by relatives or friends of the target speakers. However, this can impose restrictions on the movement and actions of the fieldworker and slow down the process. The other possibility is to check a tip on a location received from (mainly Romani) people who think they know where native speakers live, but this often proves to

be a blind alley, too.

3. **Financial limitations.** As can be seen from the first and second points, the success of fieldwork sessions is uncertain even in case of carefully selected locations and it is even more uncertain when we have to apply a kind of trial and error method: the fieldworker visits a certain location where native speakers are said to live but no speakers are found. This involves the need to travel around the country, and the travel and accommodation expenses have to be covered, as well as the fee paid to the informants. Time and money also restrict the amount of interviews that can be maid within one go.

2.11.2 The questionnaire

The questionnaire (see appendix) was specially designed by the author **to address the problems and phenomena examined in the present dissertation**. It contains 204 Hungarian sentences, which are read out to the informants, who are asked to translate them into Romani. A recording is made of the interview, which is then transcribed.

We will briefly discuss the most important aspects of the questionnaire. Several of the ideas touched upon here will be explained in detail in Chapter 6. The sentences were created by the author, in order to have more information about the weak points and variation in Lovari morphology. The three weak points to be discussed in Chapter 6 are the masculine oblique base, the feminine plural oblique base and the past tense of vocalic verbs. The sentences are read out to the informant in a jumbled order, not topic by topic.

For the masculine oblique base, I focussed on two kinds of lexical items. The first group consists of words where variation is suspected, and **they are included in the concise dictionary** of Romani dialects in Hungary (Vekerdi 2000), like *čokánó* ‘hammer’, *dúhano* ‘tobacco’, *búso* ‘bus’, *čalãdo* ‘family’, *kirãji* ‘king’, *sókro* ‘father-in-law’, *fṓro* ‘town’, *trájo* ‘life’, *pohãri* ‘glass’ etc. The second group contains **lexical items that are supposedly not part of the lexicon as such**, so informants have to provide a translation on the spot, like *laptópo* ‘laptop’, *mobílo* ‘mobile phone’, *pokrṓco* ‘blanket’, *telefóni/telefóno* ‘telephone’ etc. Here, our expectation that these words will be inflected according to the masculine paradigm proved to be right. The sentences were formed so as to contain inflected forms of the target nouns, because that is how we can see the oblique forms.

Besides the targeted items, we also acquired data about the words from the Indo-Aryan vocabulary.

For the feminine plural oblique base, **I did not differentiate 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* ‘a pair of scissors’, *patrí* ‘leaf’, *bórotva* ‘razor’; poultry and insects, like *cincə̄ri* ‘mosquito’, *tʲīrí* ‘ant’, *māčí* ‘fly’, *khajnʲí* ‘hen’, *papín* ‘goose’ etc.

The sentences were generally designed to **reflect possible every-day usage** and to sound as natural as possible. Some examples are shown in Table 2, with the target word in italics.[25]

original Hungarian sentence	English translation	Lovari translation
Elmentem a családdal a városba.	I went to town with my family.	Gēlém tar e *čalādósa* ándo fő́ro.
Leesett a kalapács feje.	The head of the hammer came off.	Tḗle pēlás e *čokanósko* šēró.
Egész nap a mobiljával szórakozik.	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* e sóba.

Table 2

Some examples of the sentences from the questionnaire, used for testing the nominal oblique base

This did not always prove to be so obvious, chiefly for two reasons. One is that something sounding very natural in Hungarian does not necessarily sound equally natural

25 It is not part of the present research, but it is important to note, that the word order of the Lovari sentences corresponds exactly to that of the Hungarian sentences. This shows one of the drawbacks of a questionnaire where the informant has to translate the sentences: we cannot know for sure whether the word order applied in the Lovari sentence is a mirror translation of the word order used in the original Hungarian sentence, or if the structure of sentences in the Lovari spoken in Hungary is generally akin (or has become akin) to the sentence structure of Hungarian. Even if this is the case, we suppose that morphology is not affected by similar effects.

in Lovari. The other one is that the fear of using words the informant does not feel authentic enough can confuse them, and these two reasons can interfere. For example, here is a sentence from the questionnaire: *Gondtól gondig tart az életünk* ‘Life lasts from problem to problem’. This may sound natural in Hungarian, but the structure of the phrase *gondtól gondig* ‘from problem to problem’ caused difficulty for most of the informants, as well as the seemingly “easy” word *tart* ‘last’.

The sentences dealing with the inflection of verbs concentrate on five aspects. The first one is the present and past tense of *-i-* stem verbs included in Vekerdi (2000), because the past tense forms vary to a great extent, and there are some signs of variation in the present tense as well. The second aspect is the past forms of *-a-* stem verbs, where there is a great deal of variation, too. The third question is the existence of additional stem-final vowels. The fourth aspect is the ways of verb derivation, as they will serve as patterns in the verb classes and tenses where inflection becomes uncertain. Finally, I also looked at the way how a novel verb is inserted into the language. Some examples are shown in Table 3, with the target words in italics.

2.11.3 The fieldwork

Within the framework of the project Variation in Romani Morphology, supported by the Hungarian Scientific Research Fund (OTKA, Project 111961, project leader: László Kálmán), we have visited and checked the places listed in Table 4 so far.

The fieldwork itself was usually carried out in the following way. Relying on either the information on Mészáros’s dialect map or personal tips, we set out to the countryside either by car or by train, taking our bicycles along. After reaching a central location in the area we wished to discover, where we had previously booked accommodation, we set out on bicycles to get to the village or villages where we were supposed to **find native speakers of Lovari**. We **made enquiries** on arriving at the village and, if we were lucky and we were not led astray, we found one or more informants. Finding the right people usually takes about one hour on average. One interview usually takes anything between half an hour and two hours, depending on how often the informant speaks Lovari or how long they have not spoken it, and also on how often we get diverted during the interview. It

can happen that an interview lasts three or four hours, because the speaker is slow or we get interrupted often.

original Hungarian sentence	English translation	Lovari translation
present and past of *-i-* stem verbs		
Segítettünk a szegényeknek.	We helped the poor.	*Žutisardám* e čorrén.
Imádkozol értem, édesanyám?	Will you pray for me, mother?	*Rudjís* pála mandé murí dej?
Levelet írtam a királynak.	I wrote a letter to the king.	Lil *iskirindém* e krajéske.
-a- stem verbs		
A lány ránevetett a fiúra.	The girl smiled at the boy.	I šej *asandás* po šāvó.
the question of additional stem-final vowels		
Nem bánok semmit.	I do not mind anything.	Či *bunúj* khančí.
Elszakadnak az ingeim.	My shirts get torn.	*Šindjón* mure gādá.
verb derivation		
Menj ki innen, mert megharagszom.	Go out of here because I will get angry.	Ža tar ārí ke *xojajváv*.
new loans		
Mindig csak mobilozik.	He is always fiddling with his mobile phone.	Ferí míndig *mobilozíj*.

Table 3

Some examples of the sentences from the questionnaire, used for testing the verbal inflection

The total number of Lovari interviews made so far is 17. We keep the anonymity of the informants; we only ask for their age, their self-confessed dialect, their place of birth, their parents' place of birth, their parents' dialect and their first name as code names for the recordings. The geographical distribution of the Lovari recordings and the age of the informants are shown in Table 5, the recordings and their transcriptions can be found in the Appendix.

county	village or town	speakers[26]
Pest	Vác, Albertirsa, Bag, Dány, Kóka	2 + 3
Komárom-Esztergom	Esztergom	2
Baranya	Pettend, Nagydobsza, Cserdi	2
Heves	Kál, Kompolt, Aldebrő, Feldebrő, Kerecsend, Mezőtárkány	0
Tolna	Dunaszentgyörgy, Fadd, Gerjen, Poroszló	1 + 3
Fejér	Sárosd, Baracska, Perkáta	1 + 2
Somogy	Balatonkiliti, Siófok	2

Table 4

Fieldwork locations visited within the project Variation in Romani Morphology

age of informants	county	village or town	number of interviews
72, 72, 82	Baranya	Pettend	3
64, 65, 67, 69	Baranya	Nagydobsza	4
58	Baranya	Cserdi	1
21, 34, 35	Komárom-Esztergom	Esztergom	3
44	Fejér	Perkáta	1
50	Somogy	Siófok	1
40, 49	Pest	Bag	2
54	Pest	Kóka	1
57	Fejér	Sárosd	1

Table 5

Lovari interviews made within the project Variation in Romani Morphology

26 0: there are no speakers. 1: there are Romani speakers but no interviews could be made due to some reason. 2: there are Lovari speakers and interviews were made. 3: there are non-Lovari speakers of Romani and interviews were made.

As we can see from the Table 5, there were three speakers under the age of 40, three between 40 and 50, four between 50 and 60, four between 60 and 70, while three of the informants were above 70. This represents a good age balance and also reflects the fact that the language is still spoken by the younger generation.

3 Analogy in grammar

In this chapter, I will introduce some core concepts of an analogy-based framework, which were taken into consideration when approaching the phenomena discussed in the present dissertation. These concepts include patterns and exemplars (3.1), similarity (3.2), synchrony and diachrony (3.3.), frequency (3.4), rich memory (3.5), paradigms (3.6), prototypes (3.7) and variation (3.8).

3.1 Patterns and exemplars versus rules and categories

Pattern generalisation is similar to, but also different from the creation of rules. Van Marle (1990) bears upon **analogy** as a synchronic morphological force and argues that 'the speakers of a language have the capacity to construct rules on the basis of the existing words' (van Marle 1990: 267), called "**rule-creating creativity**". Here, rules are taken as 'the result of a process of analysis operating on the similarity of items of the vocabulary' (Motsch 1988: 24), rather than ones that suppose abstract, underlying forms. Patterns are not so far from this definition of rules: they are based on all kinds of utterances, from sound-level to discourse-level instances that the child is exposed to, but only to truly identifiable surface elements.

For example, the Romani derivational marker *-(V)sar-* (Baló 2011) is claimed by the literature to be a complex marker, consisting of historically identifiable elements. By breaking it down into the elements *-(V)s-* and *-ar-*, we might be able to understand its origins, but, as we will see, we do not get closer to the way it actually functions in the complexity of Romani morphology. If the elements in this example are not identifiable as distinct elements, they will not serve as bases for any sort of pattern.[27]

I suppose that speakers of a language constantly process, analyse and re-analyse all the **exemplars**, from sound-level through word- and sentence-level to discourse-level. To what extent and how exactly we categorise them is not an easy question to answer. As I mentioned in the introduction, recent research into speech perception shows that we seem

27 How phonological or morphological changes can become part of the system of a particular language, so the processes more broadly known as phonologisation, morphologisation and lexicalisation (cf. e.g. Cser 2014, based on Garrett & Blevins 2009) can be interesting from a historical linguistic perspective.

to retain **very fine details** of perceptual events (Pisoni 1997: 10). On the other hand, there is a tendency in conscious human thinking to set up distinct **categories**. This instant and almost instinctive categorisation happens to everything we experience and meet in everyday life, not only in language. Linguistic categories are made up based on rules, but if rules are generalisations over patterns, how are patterns different from categories? We would like to think of categories as being distinct and having clear-cut boundaries, whereas **patterns appear to be fuzzy**. Recent findings about the cognitive processes of the brain apparently show that it creates constructions and patterns (Chandler 2002, Eddington 1996), rather than categories. Rung (2011) also adds that it is easier to build a functioning model based on the results of cognitive studies, psycho- and neurolinguistics than on the economy principle (cf. e.g. Wilder, Gärtner & Bierwisch 1997).

Let us suppose that exemplars are processed by the brain, through the application of certain, more general innate capabilities, like analogical reasoning. Figure 1 shows the possible pathway from exemplars as far as categories; patterns evolve from exemplars, rules evolve from patterns, and categories evolve from rules. If we try to find our way back to exemplars from categories, thus forming a circular process, we often find, however, that the exemplars contradict the categories we think we have found.

exemplars → patterns → rules → categories →

Figure 1

From exemplars to categories

Instead of going all the way from a vague set of exemplars to very concrete and well-defined categories, I suggest that from many different but very concrete exemplars fuzzy patterns are formed. Even if the transition goes on and our brain gets as far as rules and categories, by that stage our knowledge becomes so unfocussed that the rules and categories must be very vague, and therefore, a **gradual** approach to categories and the prototype theory is needed (Rosch & Lloyd 1978, Lakoff 1987, Langacker 1991). It also seems that the boundaries between categories, even between ones like nouns and verbs, are not necessarily so clear-cut in languages.

Following Langacker (1987) and (Bybee 2010), I argue that **grammatical patterns are accumulated sets of examples**, which can be represented as **constructions**, but the construction is not the pattern itself; rather, it is the mere **pairing of form and meaning**, a palpable representation of the pattern to be used so that we do not have to list the frequently infinite set of particular tokens. And while generalisations are indeed formed over a set of examples, the specific instances are not thrown away after the generalisation is made. This is also a core concept of a usage-based theory. Rule-based theories have relied on the idea that limitations on memory require redundancy-free representations, and any "unnecessary" information and 'particular tokens of language use' (Bybee 2010: 15) will be discarded after the generalisation is ready. In view of this, a **rule** can be defined as an **abstract generalisation** over a pattern, an **intermediate level of representation**, excluding the particular examples any longer. Considering, for example, the Hungarian plural suffix *-k*, we suppose that all the instances are stored in our memory, over which a generalisation is made, forming a pattern, which can be represented in form of a construction, a pairing of the phonological form and the semantic content. The set then constantly grows with the addition of new instances of the same pattern, and no intermediate level of representations, such as rules, is inserted.

3.2 Analogy and similarity

Patterns must be based on some sort of **similarity**, structural or functional (Itkonen 2005), which is formalised through the notion of analogy. Analogy is often defined similarly to the classic, Saussurean example (Blevins & Blevins 2009, Lahiri 2000), also called four-part or proportional analogy (cf. e. g. Kraska-Szlenk 2007): 'an analogical form is a form made on the model of one or more other forms' (Saussure 1959: 161), which he illustrates with the eventual spread of the rhotacised oblique onto the nominative in Latin. Here, the former nominative singular honōs is replaced by a more regular form, honor, as shown in (5).

(5) ōrātōrem : ōrātor = honōrem : x
x = honor

What we clearly see from this example is that analogy is often understood as a means of language change only. In the linguistic literature on Romani, we often encounter references to analogy in a diachronic sense. For example, it is often claimed that the masculine plural oblique stem (see its detailed discussion in Chapter 6) of words from Greek, like, *kókalo* 'bone' (< *κόκκαλο*) or *pétalo* 'horseshoe' (< *πέταλο*), *kokalén-* and *petalén-* changed to *kokalón-* and *petalón-* by analogy to the singular oblique *kokalós-* and *petalós-*.

Itkonen (2005) defines **four different types of relations between two analogous systems** (Itkonen 2005: 15), based on whether they exist already or not and whether they are known or not. If we think about linguistic patterns, all of the four possibilities should be considered. First, system A exists and is known, but B does not exist and is not known yet. The second and third possibilities are when B either begins to exist first or to be known first. Finally, they both exist, they are both known and they interact.

Analogy in some interpretation has been around in the study of language since the Antiquity. Structural linguists of the late 19th and early 20th centuries both in Europe and in America (Saussure, Jespersen, Sapir, Bloomfield) emphasised **the role of analogy in synchrony**, either in the creation of new words and sentences (Sapir 1921) or as the cementing force of the system (Saussure 1959). The sound laws of the Neogrammarians are revealed to have worked in an analogical way: as Phillips (2006) shows, certain types of change affect the most frequent words first and other types affect the least frequent ones first. Even what we traditionally call "contamination", so the influence of more forms in the creation of a new one (as described by Hockett 1966, for example), is a type of analogy or pattern imitation. However, the nature of and the criteria for similarity, both semantic and phonological, which appears to be essential for analogy, were not clear at all at the time. How the elements and their relations interact was also obscure.

Still sticking with a simple interpretation of analogy, it is often said to be the use of a novel item in an existing pattern (cf. Skousen 1989, Eddington 2000, Eddington 2006, Baayen 2003, Boas 2003, Krott, Baayen & Schreuder 2001, Bybee & Eddington 2006). In morphology, for example, **the productivity of a pattern or construction depends on type frequency** (Bybee 2010: 67). Although the emphasis is often placed on the novelty of the item, and in proportional analogy one form has to be missing, we are more interested in a state where there are **conflicting patterns and forms** existing already, for instance, the

vowel-zero alternation shown by a type of Hungarian verb stems, like *ugr-/ugor-* after certain suffixes (for a complete analysis in an analogy-based framework, see Rebrus & Törkenczy 2011). In case of such variation, two different sources of analogy (two different patterns) are at work. Here, the items subject to variation are not novel; **analogy is used as a synchronic tool**, as proposed in Paul (1891) already, influencing existing stems. While historical linguistics is interested in the loss of variation (analogical levelling) and the introduction of variation (analogical extension), usage-based linguistics is interested in variation itself and why variation is maintained.[28]

Let us return to **novel utterances** a little bit now. They are important in an exemplar-based model because new formations tend to rely on similarity to existing forms (Eddington 2000, Baayen 2003, Bybee 2010). English past tense forms have been a popular territory for wug-testing (Bybee & Moder 1983, Albright & Hayes 2003), especially the "irregular" past tense formations /ɪ/ – /æ/ – /ʌ/ and /ɪ/ – /ʌ/ – /ʌ/ like *ring – rang – rung* and *fling – flung – flung*. Many of these end in a velar nasal, and while nonce items ending in a velar nasal are very likely to inflect like one or the other pattern, even more interesting is the occasional behaviour of existing words like *bring* in contexts where a child's exposure to native input is temporarily less than average in a foreign language environment: as the /ɪ/ – /ɔ:/ – /ɔ:/ pattern of *bring – brought – brought* is much less frequent, the past tense of *bring* becomes *brung* in their speech. The experiments and other studies (Eddington 2000) demonstrate that not only the final consonants, but also the overall phonological similarity of the items is a determining factor, and not only that, but also the frequency of the pattern compared to other patterns, like the "regular" one in this case.

28 Analogical levelling is intra-paradigmatic and it happens when a certain stem alternant appears in a paradigmatic cell which was previously occupied by a different stem alternant. Thus, it aims at uniformity. A Polish example cited by Kraska-Szlenk (2007) is a nominal paradigm, where the earlier ʧɑs (NOM.) : ʧeɕe (LOC.) became ʧɑs (NOM.) : ʧɑɕe (LOC.) 'time', so the /e/ of the locative became identical to the nominative /a/. In other words, it happened the other way round: ʦɑna (NOM.) : ʦeɲe (LOC.) became ʦena (NOM.) : ʦeɲe (LOC.) 'price'. Although we would not like to discuss this kind of diachronic change, it is important to note that it was not a one-way process.
Analogical extension is inter-paradigmatic and introduces variation within a paradigm. Let us see the example of Kraska-Szlenk (2007) again. The original, uniform forms vizerunk (NOM.) : vizerunku (LOC.) began to alternate, and two stem variants emerged: vizerunek (NOM.) : vizerunku (LOC.) 'image', influenced by another paradigm, ranek (NOM.) : ranku (LOC.) 'morning'. Note that analogical extension also aims at uniformity, only on a different level, and it can involve the extension of "irregular" patterns, like in case of the English past tense.

Rung (2011) claims that when analogy is applied, and unusual or new forms appear, it is due to some limitation of the linguistic knowledge or competence of the speaker (Rung 2011: 19). If we consider Romani, we could say that this limitation is due to the limited use of the language, the bilingualism of the speakers, the reduced transmission of the language from one generation to the next. However, I suggest that it is the other way round: **bilingual speakers have a wider knowledge of language**, even if they do not speak their two languages as well as a monolingual speaker of any of the two. If we view it like that, this is a special, indirect effect of language contact: a possible, increased extent of analogical effects, rather than the direct borrowing or the direct use of "foreign" elements.

Krott, Baayen & Schreuder (2001), among others, show that morphology is affected by analogy, too, through the choice of linking morphemes in novel nominal compounds in Dutch. Speakers rely on the similarity between various parts of the new and the existing compounds. Even if we see that the effects of similarity are present on both the form and the meaning level, and we would like to say that they constitute a network of associations, the question **how to measure similarity** properly remains a nagging one. Especially that 'language processing seems to have a holistic component along with the more familiar linear sequencing' (Bybee 2010: 62), so for example in case of the English past tense, the whole phonological shape appears to count (Bybee 2001: 130).

Apparently, the choice of the sources of analogy is not arbitrary; we presume that there must be a relationship, perhaps mutual, between the source and the target. Therefore, it is important to select the sources with care and according to certain criteria. Analogical processing is an essential part of the structuring of language but it is also a domain-general process, which means that it operates in other areas of human cognition as well (Vosniadou & Ortony 1989, Halford & Andrews 2007, Bybee 2010).

The choice of analogical sources relies on the similarity of elements: phonemes, words, compounds, constructions (Bybee 2001, Krott 2009, Fillmore & Kay 1987, Goldberg 2006), as well as the similarity of the relations among these elements; but the further the elements are form each other, the less likely it is that the relations count (Rung 2011). The questions what factors we take into consideration when trying to grasp this similarity and what features and other properties we have to examine are more difficult ones to answer.

Establishing the factors that determine **the nature of the phonological similarity of**

words (or other kinds of similarity of other kinds of items) is an important task of any analogy-based framework. This, in turn, will help us decide which patterns should be taken into account. Similarity in grammatical function involves similarity in form, or, in other words, 'it is natural for related concepts to be designated by related sounds' (Humboldt 1999: 71).

If the calculation of similarity is done by a computer, then we must teach the computer how to do it (cf. Skousen 1989). A simple way of measuring phonological similarity is when we compare the sounds contained in two items. This is often applied in wug-testing, as in the English past tense experiments: the nonce verb *spling* (Albright & Hayes 2003) is similar to all the other English verbs ending in the sequence *-ing*, and it is also similar to those which begin with the sequence *sp-*. It is particularly similar to the existing verb *spring*, the past tense of which varies between *sprang* and *sprung*. The particular similarity is not only due to the completely identical beginning and ending sequences, but also to the similarity of the one phoneme that is different: both are liquids, often behaving phonotactically similarly.[29] This leads us to another way of measuring similarity: breaking the words down into smaller units or features, comparing them and then summing them up. When discussing the morphological phenomena in Lovari in Chapter 6, we will also rely on phonological similarity.

Linearity, as Rung (2011) points out, is crucial, so **phonological similarity is only interpreted for linearly similar strings**. Whereas *shop* and *pop* are pretty similar, *shop* and *posh* are much less similar, although the sounds are exactly the same in the second pair, whereas one sound is different in the first pair. We have to add that it would probably be wiser to take phonetic similarities into account instead of phonemic similarities, as phonemes are already based on features, that is, categories. Phonetic differences that are not considered distinctive are dismissed in a phonemic analysis, but I suggest that the phonetic properties might play a role in the mental processing instead of the phonemic ones.

29 In their experiment, in spite of the huge similarity of the verbs *spling* and *spring*, speakers were much more likely to volunteer a regular past tense form *splinged* instead of *splang* or *splung*.

3.3 Neither synchronic, nor diachronic

Structural linguists (Sapir 1921 and Bloomfield 1933) also maintained that analogy had a great significance because new utterances are created based on an analogy with previously uttered or heard words and sentences. In other words, **patterns and exemplars, already existing in our minds, serve as bases for new forms or old ones undergoing some sort of change**. It is no surprise that analogy is often linked to a diachronic approach, as opposed to a synchronic approach, which examines a state of language (Szilágyi 2011). We have to see, however, that the two approaches are not fundamentally different; both approaches are static. By taking analogy as a cementing force, through the application of which exemplars form patterns in our minds, we can also dismiss the strict dichotomy of synchrony and diachrony.

Grammaticalisation is one of the most conspicuous processes where this **dynamic nature** can be studied. It 'refers to that part of the study of language change that is concerned with such questions as how lexical items and constructions come in certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions' (Hopper & Traugott 2003: 1). Bybee *et al.* (1994) found that **there are tendencies for lexical items with similar meanings to become grammatical morphemes with similar meanings** in several languages (see also Heine & Kuteva 2002). Givón (1995) adds that 'the grammaticalisation of source domains into target domains is guided by functional similarity of potential sources and targets' (Givón 1995: 95).

Hopper & Traugott (2003), through their example *be going to/be gonna* (when English *going to* becomes the future marker *gonna*) illustrate **the impossibility of the rigidity of the distinction of the synchronic and diachronic dimensions**. As Bybee (2010) points out, 'grammaticalisation of lexical items takes place within particular constructions and further that grammaticalisation creates new constructions' (Bybee 2010: 106). The construction in which *going to* is grammaticalised is the one where a verb follows. The future construction *be going to* was facilitated by a purposive directional construction such as *I am going to London to marry Bill* (Hopper & Traugott 2003: 3) or *They are going to Windsor to see the king* (Bybee 2010: 106). The new construction is obviously different from the old one, as the verb *go* has lost the sense of movement. Based on Bybee & Pagliuca (1987) and Bybee *et al.* (1994), among others, we can say that

grammaticalisation is a product of language use. Bybee (2010) claims that it 'involves the process by which a particular lexical instance of a construction (*go* in the purpose construction) becomes autonomous from the other instances of the construction' (Bybee 2010: 107), and that analysability is lost. I would argue that analysability is lost indeed, but it is not just the lexical instance *go* that takes part in the process: it is the more complex form *be going to*, which, exactly through the change of meaning, comes to be an individual entity that cannot be parsed or analysed (cf. the phonetic reduction to *be gonna*). **The grammaticalisation process induces change in the frequency of the use of the construction**, which, in turn, influences the applicability of other forms with a similar content, like, in this case, the future marker *will*. In every-day speech, *be going to* can replace *will* in constructions where only *will* was possible, like the conditional. Bybee (2010) adds that studies of current variation and ongoing changes reveal that 'we do not find gaps or abrupt changes across generations … but rather that even fairly young children produce variants of linguistic forms that are good reflections of the adult variation' (Bybee 2010: 116-117), which is exactly what a usage-based approach predicts. **As a construction becomes more frequent, the more likely it is that it will spread through the language use of adults**, especially if we presume that adults can generalise constructions more creatively. Haspelmath (1998) also emphasises the gradual nature of grammaticalisation as opposed to abrupt reanalysis. Thus, 'the gradient facts of usage, synchronic variation and gradual diachronic change could be taken as principal evidence that grammars themselves incorporate the gradience and variability' (Bybee 2010: 120).

A Hungarian example is the spread of the construction [ADVERBIAL + *hogy* 'that' CONJ. + CLAUSE] perhaps instead of either [ADVERBIAL + CLAUSE] or [ADJECTIVE + *hogy* 'that' CONJ. + CLAUSE]: *valószínűleg/természetesen/remélhetőleg, hogy*, perhaps instead of *valószínűleg or valószínű, hogy*. É. Kiss (2010) shows that the phenomenon is not new at all (there are data from as early as the 19th century), and that its use and acceptance does not depend on the social status of the speaker (Kontra 2001). Although É. Kiss (2010) tries to explain the emergence of the construction by the contact influence of Romanian, the point is not where a change comes from, because change itself is an internal and inherent property of language.

By this, we also mean that neither changes, nor states can be described discretely, like synchronic (generative and descriptive) and diachronic (historical) linguistic studies

suggest. **Language is a constant interaction among its elements** (Itkonen 2005). In that sense, language in general is similar to the the language of an individual, which is never static either (Bybee 2010). A good argument comes from Rung (2011), who says that while it is possible for a given state to contain data that are historically distant, through, for example, the use of idioms, quotes, archaisms (Rung 2011: 10), there can be phenomena which are processed over a very short period of time.

3.4 Variation and frequency

Variation, the role of frequency and the gradual nature of linguistic phenomena mentioned in section 3.3 can be accounted for in analogical framework, too. Even if crucial information is missing due to external factors (Skousen 1989: 58), an analogical approach can predict the outcome quite reliably: if the first sound of a word is blocked out by some noise, for instance, and the listener cannot hear it, it can still predict which allomorph of the English indefinite article will be used, based on the segments following the missing sound. **An analogical approach is by definition usage-based**: it has got nothing else to resort to but actual language use. A usage-based approach (Bybee 2010), in turn, emphasises the psychological and social functions of communication.

In a framework based on language use and analogy, the difference between regular and irregular loses its significance, too. The same processes govern the formation of regular patterns and the more gradient, variable and less frequent patterns. Their significance is therefore equal. Discrete categories only exist in synchronic and diachronic terms, where language is seen as a fixed structure; but language is in constant use.

Gradience and variation can come in many shapes and in all fields of language and morphological categories are no exception. Bybee (2010) presents a few examples of gradience in variation in linguistic structure (Bybee 2010: 2-6). **There is no strict boundary between lexical and grammatical morphemes**, as the examples of *go* shows. Apart from its strictly lexical use, it appears in many other constructions, more like a grammatical morpheme, such as in *go wrong*, *go see who it is*, *I am going to do it*, *and then he goes "sod off"* etc. This kind of change has been around in language (instances of it have been noted in Romani, too, cf. e.g. Friedman 1991a, b on postpositions becoming

case markers), but it has only been observed as a change, a transition from one category to another, or as the morphological recycling (Booij 2008) of items and not as something that depends on the construction the morpheme takes part in.

> Grammatical morphemes are classically defined as closed class items. Since classes are defined in terms of the properties of constructions, grammatical morphemes are those which are restricted to particular positions in constructions. (Bybee 2010: 3)

On the one hand, if *go* can become a grammatical morpheme (and we cannot fail to notice the functional similarity of *go* to English auxiliaries like *be*, *have* and *do*), then grammatical morphemes do not constitute a closed class. We could say that the grammatical *go* and the lexical *go* are different morphemes, but this would have no foundation. On the other hand, if we consider *go* one morpheme, then the restriction to particular positions is not valid either.

How usage factors influence variation is shown by Bybee & Scheibman (1999) through the negated auxiliary *do*. They look at spoken American English conversation and examine the possible contexts where *don't* can appear and find that phonological reduction appears in more frequent contexts. According to the author's observation, we can find its equivalent in British English, too: the full form /dəʊnt/ appears with less frequently used verbs, whereas such frequent verbs as *know*, *care*, *think* induce its reduction to [ʔə̃] in sentences like *I don't know/care/think*.

The role of frequency has also been emphasised since the beginning of modern linguistics (Wheeler 1887), as something that strengthens a given element. The lack of corpora at the time, however, did not make it possible to find out more about that. Unfortunately, in the case of the Lovari dialect of Romani, the lack of an appropriate corpus (or a corpus as such) still makes it very difficult to say anything specific about frequency effects.

Again related to English past tense forms, Chandler (2002) notes that frequency might play a role in opposition to similarity, so the two forces, similarity and frequency might go against each other, when a more frequent pattern is chosen over a more similar one. Anyhow, frequency, or, in other words, the robustness of a pattern does have an effect on the language. Frequency can refer to at least two different things: token frequency and type

frequency.

> 1. Items with high token frequency have greater lexical strength and therefore resist morphological or analogical change, serve as the basis for change, and have greater autonomy.
> 2. Patterns with high type frequency are more productive than patterns with low type frequency.
> 3. As high token frequency leads to greater autonomy, items with high token frequency have weaker connections to related forms and thus are more likely to become independent and less likely to contribute to the formation of productive classes. (Bybee 2001: 136)

The example of the reduction of *don't* quoted above supports the suggestion that more frequent forms or constructions are more prone to phonological change (Bybee & Scheibman 1999). An appealing property of frequency is that it is more easily measurable than similarity. Frequency might also play a role in pattern conflict or variation, where the bigger dominance of a pattern might be due to either type or token frequency. As it is argued by Wedel (2007, 2009), similarity-biased errors and positive feedback can be instrumental in the development of consistent, dominant phonological patterns. Referring to Hare & Elman (1995), Wedel (2009) mentions that the same thing is true for morphological patterns. What is interesting for us is the claim that uniformity cannot, by default be maintained overall the system. If a subsystem starts to become more organised, another one might begin to become disorganised.

3.5 Rich memory and exemplars

A usage-based, analogical approach heavily relies on exemplars. **An exemplar is not an individual item, it is rather a set of identical tokens.** The frequency of the exemplar can be measured by way of the number of the individual occurrences. Although the role of memory has been questionable in children's language acquisition, it would be difficult to question that any individual occurrence of the same token reinforces the role and importance of that exemplar. Similarly, encountering new tokens will lead to the

rearrangement of the existing patterns and influences, as ‘structure emerges from use’ (Tomasello 2009: 69). According to Langacker (1987), **after reaching a generalisation, the items on which the generalisation was made do not have to be thrown away** – and they are probably not deleted from our memory. This makes it possible for speakers to rearrange these items and the generalisations when a new, perhaps contradictory piece of information is heard. In Hungarian, the forms *látszódik* and *hallatszódik* are increasingly popular instead of *látszik* and *hallatszik.* If certain linguistic information becomes obscure (the mediopassive nature of *látszik* and *hallatszik*), then the function is reinforced and marked more saliently. Similarly, the more one hears the unusual form **csukoljon*, the more it is likely to become a possible form (cf. also Grétsy & Kemény 2005), as the individual occurrences get stored in the memory.

Our memory contains a plethora of linguistic information, including redundant and variable items and constructions, and this is exactly what recent research points to (Bybee 2010). **Redundancy is part of the system, and ‘analogy is one essential way in which redundancy can be discovered by the language learner’** (Goldsmith 2009: 149). The rich memory model is often dismissed partly on the grounds that human memory is limited and we simply cannot remember such a huge amount of input, partly because it is hard to believe that a complex system like language can be learnt by means of such a primitive thing as imitation.[30] If we think of the process of child language acquisition in terms of the rich memory and imitation model, then we might be able to explain it. Children start to speak for the first time when they have gathered enough input from which they can start to generalise, and they go through a long phase of trial and error. If we relied on innatism, there would be no need for the trial and error phase to be so long, and if the frequency of items did not count, children could speak fluently and “properly” much earlier: they register that something is said in a particular way in their mother tongue and then they should immediately know which part of Universal Grammar they should use and ignore the rest of Universal Grammar concerning the same item or grammatical aspect. It is more plausible to think that every time a token or type is heard, it strengthens the validity of the given item or construction. Thus, exemplars are complex sets of tokens. As for phonology, Bybee (2010) claims, referring to Pierrehumbert (2002) and Bybee (2006), the following.

30 Cf. the poverty of the stimulus (Chomsky 1980).

> …each of the phonetic forms of a word that are distinguishable are established in memory as exemplars; new tokens of experience that are the same as some existing exemplars are mapped on to it, strengthening it. Then all the phonetic exemplars of a word are grouped together in an exemplar cluster which is associated with the meanings of the word and the contexts in which it has been used, which themselves form an exemplar cluster. (Bybee 2010: 19)

This process works the other way round during language production. Once the speaker has got the meaning to be conveyed, the phonological exemplar cluster is retrieved, and then the strongest phonetic exemplar.

According to Bybee (1985, 2010), **variation in morphology can also be a result of different sets of phonetic and semantic exemplars**. In her view, the similarity of the final consonant of verbs like *play*, *spill*, *spoil* etc. (and the fact that they contain other, identical consonants, like /p/), in addition to the past tense meaning, so shared phonetic and semantic features lead to the emergence of forms like *spilled* and *spoiled.* We have to add, though, that these forms, as opposed to forms like *spilt* and *spoilt*, can be attributed to more general tendencies. Bybee (2010) herself also mentions that *had* may also be part of the cluster of semantic and phonetic exemplars, with its final /d/ and past reference, and that 'one advantage of this approach to morphological analysis is that it does not require that a word be exhaustively analysed into morphemes' (Bybee 2010: 23), as in the Lovari example mentioned in Section 3.1.

The strength of exemplars also depends on **a network of associations and relative frequency**. For example, Hay (2001) looked at the level of complexity of complex words. She experimented with words like *happiness*, which is very complex, and *business*, which is less complex, while both can be broken down into *happy* + *-ness* and *busy* + *-ness*, respectively (Hay 2001: 1048). Similarly, there were other prefixes and suffixes involved in the experiment. While Hay (2001) examines 'if words that are more frequent than their embedded bases appear more easily decomposable than words that are not more frequent than their embedded bases' (Hay 2001: 1046), the experiment can also be used to show that the strength and productivity of an affix depend on the relative frequency of words that seem to contain it and the "base" word, as well as the semantic relations, e.g. *dismiss* and *miss* or *canny* and *uncanny*, and all this can be handled by a rich memory approach.

An approach like that obviously relies on '**closely related surface forms**' (Bybee 1985: 49-50, and see also e.g. Kálmán, Rebrus & Törkenczy 2011). There is no need for segmentation, because the behaviour of surface forms is not determined by rules that define the combination of some elements, but on the relationship of the surface forms (Ackerman, Blevins & Malouf 2009). The same is true for word formation, where 'patterns can be seen as abstract schemas that generalize over sets of existing complex words' (Booij 2007: 34).

Syntactic constructions as defined, for example, by Goldberg (1995, 2006), as pairings of form and function, similarly to the notion of sign taken in the original, Saussurean sense: 'the linguistic sign unites ... a concept and a sound-image' (Saussure 1959: 66), are also instances of exemplars. The individual tokens experienced by language users can represent and relate to several different constructions. **Analogical relations exist among constructions**, and the reapplication of the same construction with different elements is analogous as well.

The idiomatic resultative construction studied by Boas (2003) (and also quoted by Bybee 2010), *it drives me mad*, for example, is part of a complex set of relations. First and foremost, we have the narrowest interpretation [SUBJ. *drive* OBJ. 'mentally ill'], but even this is subject to frequency effects as to how often different tokens (*it*, *you*, *my wife/husband*, *your mother*, *my neighbour* etc.) fill the gaps of the subject, the object (*me*, *my wife/husband*, *her* etc.) and the semantic content 'mentally ill' (*mad*, *crazy*, *insane* etc.). A google search of the strings "drives me mad", "drives me crazy" and "drives me insane" shows a more or less equal number of results (169, 142 and 153, respectively), but if we start changing the variables, the results will probably differ more significantly. All the different instances have an an impact on the construction as an exemplar, and similar representations of other constructions have an impact, too. For example, as the instance *he always drives me mad* is similar to the representation of another construction [SUBJ. *drive* OBJ. 'location'], *he always drives me to the railway station*, their strength and frequency mutually influence those of each other.

3.6 Paradigms

Paradigms, that is, a set of forms belonging to the same lexeme (cf. Wurzel 1989: 52), play an important role both in the phenomena observed in Lovari morphology and in analogy-based approaches in general, as the similarity of combinations of form and function is a significant characteristic of paradigms and because one of the main criteria for belonging to a certain paradigm is based on some kind of similarity of the items (Albright 2009), and paradigm uniformity (Steriade 2000), that is, when even a non-contrastive phonetic feature of the base of the paradigm can be retained all over the paradigm, is also maintained by analogy (Eddington 2006). Two different morphological models have been around, the **"constructive" morpheme-based approach**, going back to Pāṇini and his Sanskrit grammar (no surprise that this approach so frequently appears in the literature dealing with Romani) and the **"abstractive" word-based approach**, going back to Aristotle (Blevins 2006 and 2013). We will adopt the **surface-oriented definition of a paradigm** as formulated by Rebrus & Törkenczy (2011): it is '**a set of word-forms associated with the same morpho-syntactic function**' (Rebrus & Törkenczy 2011: 126). A word-based approach works better because 'it is often the case that larger units unambiguously predict smaller units, whereas the smaller units are of more limited predictive value' (Blevins 2006: 568). It can deal with multi-purpose and cumulative morphs more easily by simply not dealing with the level of morphs. Stump & Finkel (2013) describe a word and paradigm model (Bauer 1988: 196-213) by relying on **principal parts**.

> For complete clarity, we define the paradigm of a lexeme L as a complete set of cells for L, where each cell is the pairing of L with a complete and coherent morphosyntactic property set for L for which L is inflectable. … The principal parts of a lexeme L are a set of cells in L's realised paradigm from which one can reliably deduce the remaining cells in L's realised paradigm. (Stump & Finkel 2013: 9-11)

Stump & Finkel (2013) differentiate among three schemes for optimal principal-part sets (Stump & Finkel 2013: 29-35), all of which can be useful. In the static scheme, the principal-part specification of a lexeme L is uniform across the inflection classes. In the adaptive scheme, the first principal part is uniform across the inflection classes, while the

other principal parts can vary. In the dynamic scheme, the principal parts are neither linearly ordered, nor necessarily uniform from one inflection class to the other. They also introduce the notion of maximal transparency and maximal opacity for inflection classes. An inflection class is maximally transparent, if each individual cell can determine the other cells in all its realised paradigms, while it is maximally opaque, if none of the cells of any realised paradigm can determine any other cell. The more transparent an inflection class, the more it is determined by analogy. It would be interesting to know how frequency influences the possible principal parts of a paradigm.

An alternative word-based analysis is presented by Ackerman & Malouf (2016), where they rely on **more complex implicative relations within paradigms** instead of picking out particular cells to serve as base forms, or in the terminology of Stump & Finkel (2013), principal parts. Ackerman & Malouf (2016) refer to Ackerman *et al.* (2009) when they formulate the question as the Paradigm Cell Filling Problem: 'given exposure to an inflected wordform of a novel lexeme, what licenses reliable inferences about the other wordforms in its inflectional family?' In addition, they point out an important fact which is very relevant to Romani, namely that 'in any inflectional system, some classes will have more members than others, and a randomly selected lexeme is more likely to be a member of a class with many members' (Ackerman & Malouf 2016: 16). Type frequency is the number of lexemes that are members of a class, and the probability of a randomly selected word being in an inflection class is the number of lexemes of that inflection class divided by the sum of the lexemes of all inflection classes. The solution to the paradigm cell filling problem will also depend on cell probability, the probability of a randomly selected wordform being the realisation of a cell, which in turn depends on token frequency.

3.7 Patterns and prototypes

Analogical models, instead of underlying forms and morpheme-based segmentation, usually work with **carefully selected patterns of surface forms, based on similarity relations**. While exemplar representations should 'contain, at least potentially, all the information a language user can perceive in a linguistic experience' (Bybee 2010: 14), it is often impossible to extract even small amounts of phonetic aspects from a corpus, let alone

phonological properties of exhaustive phonetic detail, although they do seem to play a role (cf. Bybee & Hopper 2001, Cohn *et al.* 2012).

> …it is not possible to point to a single case in which analogous phonemes in two different languages display exactly the same phonetic targets and the same pattern of phonetic variation in different contexts. Exact phonetic targets and patterns of variation must accordingly be learned during the course of language acquisition. The usage-based framework readily accommodates such findings by proposing that mental representations of phonological targets and patterns are gradually built up through experience with speech. (Pierrehumbert 2001: 137)

The phonetic-phonological interface is just one aspect where, at least at the moment, we have to face a deficit of the analogical approach and analogical modelling. In a corpus-based analysis, we have written, often abstract interpretations of what is said and heard actually. Once we have contented ourselves with those, there are still questions concerning what specific words, patterns or groups of surface forms (paradigms) or prototypes exert influence in a particular case.

The lexical strength of an individual item is also important. As Bybee (1995) claims, 'words with high lexical strength are easy to access, serves as the bases of morphological relations and exhibit an autonomy that makes them resistant to change and prone to semantic independence' (Bybee 1995: 428). **Lexical strength depends on token frequency**, and that is why we see that irregular forms which are more frequent are more stable at the same time, while less frequent irregular forms are less stable and more easily influenced by a more regular pattern. Thus, less frequent patterns survive because either the pattern itself is not frequent but the individual items exhibiting the pattern are; or because they are phonologically similar to other, more frequent items.

For example, we may presume that a Lovari word like *főro* 'town' keeps its more irregular oblique form *fōrós-* because of its high token frequency, while the oblique form of a word like *čókano* 'hammer' varies between the more regular *čokanés-* and the less regular *čokanós-* because of its lower token frequency. **More generally, we can also suggest that paradigms have token frequency, too, which is the sum of the token frequencies of all the members of the given paradigm.** If we want to use the oblique form *fōrós-* of *főro*, we might have immediate mental access to it. In case of other words,

like *čókano*, where either the oblique form, or the word itself, or the whole paradigm, or all three are less frequent, we have less access to it, and it begins to vary between the more and less regular variants. Alternatively, it is also possible that both forms are stored in the memory, characterised by different frequencies. When we have to produce the form, we either produce the form whose lexical strength is higher, or produce the one with lower lexical strength because other factors, like the expectations of the listener, interfere, too. If no forms are stored in the memory, as it happens in the case of on-the-spot borrowings in Lovari, token frequency will not play a role: *mobilozíj* 'he/she plays/talks on his/her mobile phone' will be under the influence of the gang effects or type frequency of different verb classes.[31]

We might also add, following Rung (2011), that the probability of the occurrence of a certain form depends on the constantly and dynamically changing frequencies of similar forms and their relations. It is a function of the whole, complex system of language, but the more distant another form is, the less likely it is that this form exerts any influence (this is the degree of relatedness, cf. Bybee 1985). Let us take the word *dúhano* 'tobacco'. As we will see in Chapter 6, the probability of the oblique forms *duhanés-* and *duhanós-* seem to be equal based on the data we collected. This is the result of several interrelated factors. One of them is the frequency of the related forms (in other words, paradigms) of words ending in the sequence *-ano*, like *čókano* or *romanó* 'Romani'. This is, however, complicated by the fact that *čókano* itself also alternates. The distance of categories (or syntactic positions) must count, too: the adjective *romanó* and the noun *čókano a*ppear in different positions in a sentence. The effect of *romanó*, on the other hand, is enhanced by the common phrase with high token frequency, *žanés romanés?* 'do you speak Romani?'. Words ending in the sequence *-uno*, like *zúbuno* 'coat', or *-ino*, like *cinó* 'small', being less similar to *dúhano*, will have less effect, but will still be of some influence. If *zúbuno* alternates (which seems plausible but we do not know because we do not have enough data to tell for certain), it will create further disturbances. As phonological similarity must be a factor, all other words ending in *-o* will have en effect, but considerably less than words ending in *-no*. Obviously, wordforms ending in *-es* and *-os* will also be influencing factors, probably more if they are oblique forms and less if they are more distant category-wise.

The dynamic nature of language also implies that **these relations interact**, so while

31 Rule-based descriptions of Romani explain the adaptation of borrowings into a certain verb class by positing a set of complex rules of derivation, suffixation and deletion of suffixes.

we focussed on the word *dúhano* just now, the same effects are valid for *čókano*, *romanó*, *zúbuno*, *cinó* etc. These create a complex network, similar to the Network Model described by Bybee (1998). Similarity and frequency effects also interact in this network. When looking for the oblique form of *dúhano*, we recall the forms that are the most similar and the most frequent at the same time (Jurafsky 2003). This means that not only the patterns are in competition, but also frequency and similarity, which explains why it is not straightforward which pattern will win: sometimes it is the one that is more frequent but less similar, at other times it the one that is more similar but less frequent.[32] Bybee (1985) already emphasises that the 'productivity of morphological rules must be connected to high type frequency' (Bybee 1985: 133), later repeated in Bybee (2001) and Bybee (2010), among others, and this suggestion has been reinforced by other studies (Baayen 1992, Eddington 2003, Albright 2009); in addition, the Lovari case also seems to corroborate the idea. On the other hand, Bybee (2010) emphasises the conserving effect of high token frequency (they are less likely to undergo change, Bybee 2010: 75), and, naturally, items of high token frequency can become prototypical items.

The notion of prototypes has been around since the emergence of the Prototype Theory (Rosch & Mervis 1975), which was further investigated by Lakoff 1982 and Langacker 1987, 1991, among others. Taylor (2003) gives a thorough overview of prototypes (Taylor 2003: 38-80). The notion of prototypes as defined by Langacker (1987) is the most relevant for an analogical approach and is closely related to the notion of schemas. In his view, '**a prototype is a typical instance of a category**, and other elements are assimilated to the category on the basis of their perceived resemblance to the prototype … **a schema, by contrast, is an abstract characterisation** that is fully compatible with all the members of the category it defines' (Langacker 1987: 371). Categorisation based on prototypes corresponds to the idea I introduced above concerning the relative distance of all the forms in a language[33] from the form in question. Schemas are also important as they represent groups like declensions and conjugations which are ostensibly different from

32 Contrary to nearest neighbour models (e.g. Aha, Kibler, & Alber 1991 and Cost & Salzberg 1993), which suggest that similarity is more important, in an experiment for English past tense forms, Chandler (1998) found that subjects preferred patterns with high frequency, even if they were less similar (cf. Chandler 2002). This also implies that rare forms rarely serve as analogical patterns (Rung 2011). However, highly suppletive forms, like *went*, are unlikely to be bases of nonce verbs, in spite of their high frequency.

33 If we approach it from the point of view of cognitive sciences, areas other than language and mental processes other than linguistic might be considered, too, which might influence the choice of the form.

each other. We have to emphasise again that neither prototypes nor schemas need to be determined very precisely, as a lot of information could be lost. We can say that prototypes are based on frequency and schemas are based on similarity, but they are both two ways of saying the same things.

3.8 Variation

Under the term **variation** we mean that at any given time, **at least two different forms can fulfil the same grammatical function without any grammatical or other reason**. Although variation of this kind can be the result of several things, we will only consider it as "true" variation if the appearance of form A or B does not depend on any external factors such as the following:

a) the varieties spoken by the speakers who produce variation belong to a different dialect;
b) the varieties spoken by the speakers who produce variation belong to a different sociolect;
c) the observer's paradox.

Factor a) is an especially sensitive question in the case of Romani, as there is a high number of dialects and dialect branches. When trying to exclude it, the researcher must rely partly on the self-designation of the informants, partly on their own knowledge of the dialect. The variety discussed in the present paper, commonly called Lovari, is a fairly well-documented dialect, although regional differences may occur. However, if most of the features which are not under scrutiny are identical, the researcher can be confident enough that **the variety spoken by the different informants is one and the same**.

As for factor b), we have to be aware of the fact that the language itself currently constitutes a form of sociolinguistic layer, being almost exclusively a second language. Speakers often say that one of the purposes of avoiding the use of borrowings from Hungarian is to make sure that people who might overhear the conversation but do not speak Lovari should not understand what they say. Szalai (2015) suggests that there must be different styles, as well as formal and informal registers within Romani. When we

collected the data, the circumstances were generally the same, so **these aspects did not influence the instances of variation**.

The term observer's paradox was coined by Labov (1972) and it refers to the problems that arise when, under the influence of the presence of the fieldworker who wishes to observe the way the language is naturally spoken, speakers start to speak differently, usually trying to adopt a variety with a higher prestige. However, even if there are varieties with higher and lower prestiges within Lovari, **all the data are gathered in exactly the same way, so even if observation influences the way people talk, it does so in each and every case**. On the other hand, the observer's paradox presents itself in another form: a Romani-speaking informant might feel "ashamed" if they use too many Hungarian loanwords, albeit inflected in accordance with Romani morphology; therefore, when facing a sentence in which they would have to use many Hungarian words, they often fail to give an answer on the grounds that "it would be in Hungarian".

Any grammatical system at any time is unstable, and this instability manifests itself in variation. **Variation is often linked to change**, as linguistic changes are preceded by a stage where a certain morphological item, having form A, begins to appear in form B. For example, less typical elements on the boundaries of categories (Chandler 2002) may be more prone to undergo change. However, it is also true that change precedes variation: some kind of change has to take place which can trigger variation, which in turn triggers a more fundamental change. To take Saussure's classic example again, Latin rhotacism (the change of intervocalic /s/ to /r/, like in *honōsem* > *honōrem*) triggered the variation and eventual change of *honōs* to *honor* (Saussure 1959, Davies 2004). This can also be viewed as **a constant struggle for regularity within a system where apparently unjustified changes push it towards irregularity**. Changes, however, are not necessarily unjustified. Regularisation on one level of the system may result in the appearance of irregularity on another, or, in other words, simplification of one aspect of the language can go hand in hand with another aspect becoming more complicated. While the aim might be to reach a more stable state and to get rid of forms that do not "fit", the process can create forms which do not fit some other part of the system; therefore, the system can never reach an ideal state.

According to the "written language bias", as outlined by Linell (2005), **linguistic models and theories have depended on written language**, and there is still a strong bias

in its favour within mainstream linguistics. (Linell 2005: 37). An assumption reflecting this written language bias concerns the unity and homogeneity of each language and includes that 'variation is not a property of the language system *per se*', but we must see that 'no known languages are devoid of sociolectal, dialectal or idiolectal variation' (Linell 2005: 45-46). Variation, neglected for some time, has recently regained interest (for example the nature of and the possible factors playing a role in "vacillation" as seen in vowel harmony in Hungarian, cf. Törkenczy, Szigetvári & Rebrus 2013).

Variation is perhaps most conspicuous when forms A and B are used by the same speaker, and the number of speakers using both forms is high. Speakers constitute a central part of the system, as, although 'a change is observed within the language system' (Milroy 1992: 169), that is where innovations originate from. Of course 'if a child creates a neologism they have not changed the language' (Allen 1995: 15).

> It is in speaking that the germ of all change is found. Each change is launched by a certain number of individuals before accepted for general use. (Saussure 1959: 98)

When discussing variation, it is worth just briefly looking at the notion of **speech errors**. On the one hand, generative linguistics originally dismissed errors, counter-posing competence to performance.

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965: 3)

However, by focussing on an ideal speaker-listener, we might lose important insights into language use by overlooking variation. On the other hand, the sociolinguistic school sometimes places perhaps too much emphasis on errors.

> As a whole, the distinction between competence and performance has never proved to be particularly helpful for our work, and it becomes more and more unclear as

> we consider its general implications. If performance involves limitations of memory, attention, and articulation, then we must consider the entire English grammar to be a matter of performance. (Labov 1971: 468)

The truth could be somewhere in between: arbitrary speech errors, hesitations and attention gaps might reveal a lot, but variation is more significant because **it is not random and unpredictable**, like speech errors are. It can be the result of several external (language contact) and internal (for example errors) factors.

Sometimes it is even said that 'change *is* variation' (Labov 1982: 20), and variation is indeed a remarkable sign of the dynamic nature of language. It involves the emergence of different patterns, a natural phenomenon in languages. Any language is composed of a complicated network of intertwining patterns which are constantly prone to change. The task of linguistics is not to find all these patterns; it is impossible to find them as there are so many, they often overlap, and this or that aspect might change by the time we find an appropriate way to describe them. And when there are conflicting patterns, we might expect that one of them will eventually become prevalent (Wedel 2009). But even if some sort of a balance is reached within one part of the system, there will be other, weak parts where it will become unbalanced. The general view is that less frequent forms become similar to more frequent ones, and items which do not fit into a category or paradigm try to change in order to fit.

Part Two

4 The Lovari sound system

4.1 Consonants

	bilabial			labiodental		dental			alveolar		postalveolar			palatal		velar			laryngeal
stop	p	p^h	b			t	t^h	d						c	ɟ	k	k^h	g	
fricative				f	v				s	z	ʃ		ʒ		ʝ	x			h
nasal			m					n							ɲ				
trill								r											
lateral								l							ʎ				
affricate									t͡s		t͡ʃ	t͡ʃʰ							

Table 6

Lovari consonants

The dental stops, the dental nasal and the lateral liquid /d, t, n, l/ are separate phonemes from the palatal stops, the palatal nasal and the palatal liquid. In keeping with the traditional transcription and for the sake of orthographic ease, the latter ones will be transcribed throughout the text as /t^j, d^j, n^j, l^j/. The aspirated non-palatal voiceless stops /p^h, t^h, k^h/are also separate phonemes. The palatal fricative appears as a glide in a syllable-final position. As yet it is undetermined whether this is a territorial difference or something else, but certain speakers produce a palatal stop where others have a velar stop, for example *kiro-* versus *t^jiro-* ‘cook’ INTR. Length is not a distinctive feature of Lovari consonants, although geminate consonants do exist, especially for the palatalised dental stop: *katt^{j}a(n)-* ‘scissors’ OBL. and the liquids, although it seems to be optional there: *cerra* ‘little, few’, both /t͡sera/ and /t͡ser:a/.

4.2 Vowels

	front	back
	unrounded	rounded
close	/i/	/u/
mid	/e/	/o/
open	/a/	

Table 7

Lovari vowels

Although Lovari vowels can be long or short, length does not play a distinctive role, therefore we will not deal with it here. Lovari is very conservative in that it retained the basic, Early Romani vowel inventory /a, e, i, o, u/. The front rounded vowels /y/ and /ø/ from Hungarian loanwords are always adapted and become unrounded.

4.3 Stress

According to Hutterer & Mészáros (1967), the stress in Hungarian Lovari is essentially stem-final. Matras (2002) claims something similar: ‘stress is on the final position of lexical roots’ (Matras 2002: 63). In addition, both of them remark that certain affixes do not attract stress, so for example *bakró* ‘sheep’, *bakrés* ‘sheep’ ACC. and *bakréske* ‘sheep’ DAT. The root- or stem-final stress, according to them, extends to the whole morphology. Matras (2002) adds that adverbs like *anglál* ‘in front’ also bear word-final stress, but European loanwords retain the original, usually word-initial stress in the nominative (*fő́ro* ‘town’ *forósk*e ‘town’ DAT.), while Hutterer & Mészáros (1967) claim that the stress of loanwords vary (*bárāto* and *barātó* ‘friend’). According to the observations of Endre Tálos (personal communication) and the author, however, this is not quite the case.
Firstly, penultimate stress is not uncommon at all: *lulúdʲi* ‘flower’ (besides *ludʲí*), *paramíči* ‘tale’ (besides *paramičí*), *vonáto* ‘train’, *hajő́vo* ‘ship’. Secondly, the “original” stress is not necessarily retained: *simadʲí* ‘pawn’ (from Greek *σημάδι*), *vonáto* ‘train’ (from

Hungarian *vónat*).[34] Based on these data, we cannot say that stress in Lovari, or Romani on the whole would be stem- or root-final. Rather, we might suggest that there are different stress patterns (word-final, stem-final, penultimate, word-initial etc.), which influence each other. Forms like *bakréske*, where surface stress falls on the penultimate syllable, can influence other, morphologically unrelated forms, such as nominative forms like *lulud^{j}í*, which results in variation in stress.[35]

34 If Romani had a uniform stress pattern, stress assignment for loanwords would be more straightforward. However, as we can see, this is not the case. Loanwords either retain their stress or not; for example, Hungarian loan nouns often receive penultimate stress, which is more similar to the stress pattern of Romani inflected nouns. As stress varies on the surface (word-final for nouns in the nominative and accusative but penultimate for inflected forms), it does not come as a surprise that the stress of loanwords will not follow one particular pattern, although there seems to be a tendency to lose, rather than to retain its original stress, and to acquire more "inherent" stress patterns (see e.g. the case of English loanwords in American Hungarian, where word-initial stress is dominant and English words are adapted to the word-initial stress pattern, Fenyvesi & Zsigri 2001).

35 An interesting case is the stress of mediopassive verbs (to be discussed in detail in section 5.2.2.2). According to Endre Tálos (personal communication), verbs like *bə̃ro-* 'grow' receive their stress like this: the stress shifts to the front position when *bāró* 'big' + *ov-* 'become' (or *-uv-*, a derivational marker) becomes *bə̃rov-/bə̃ruv-*. Thus, for example, the first person singular past indicative form is *bə̃rilem*. However, *bārilém* also exists, which suggests again that the stress patterns are not fixed morphologically and are under the influence of each other.

5 Noun classes and verb classes in Lovari

An excellent area for testing an analogy-based framework is inflectional morphology. In the present dissertation, we will focus on **instances of vacillation in the nominal and verbal inflection of Lovari and how we can possibly explain them by means of analogy**. In order to see the processes clearly, we will first have to **establish how many noun and verb classes there are in Lovari** after all, what these are, and what features we can use to separate them. This will also involve a critical overview of the literature on both the declension and the conjugation of Lovari.[36]

5.1 The nominal inflection

In order to comprehend the possible analogical forces at work in the morphology of Lovari, we have to be familiar with the nominal inflection of Lovari. Grammars and linguistic descriptions of Romani, Vlax Romani and Lovari in particular (for example Hutterer & Mészáros 1967, Matras 2002, Elšík 2000 specifically about the nominal paradigms, Cech & Heinschink 1999, Boretzky 1994, Boretzky 2003) mention several definitive properties for nouns, presented from either a descriptive or a historical perspective. In this chapter, I will introduce, clarify and dismiss these features where necessary. The main features we will go through are the following.

1. Gender. In connection with gender, we will see that although it is sometimes deducible from certain formal or semantic features, not surprisingly, it is rather arbitrary otherwise, and, under certain circumstances, it can even be neutralised in the plural, as we will see in section 6.4. Ultimately, however, we will also see that **this feature defines the two noun classes in Lovari**. Two of the three weak points in Lovari morphology, to be discussed in the following chapter, are separated and connected along grammatical gender. The first weak point to be seen in the oblique form of the masculine paradigm, while the second one is present in the plural oblique form of the feminine paradigm, where the variation also

36 Henceforward, I will use the term Romani when referring to information that is generally valid in the language, and the more specific name Lovari when talking about phenomena which are only typical of Lovari.

involves an instance of neutralisation between the masculine and the feminine. The subclasses of nouns, sometimes posited based on other features, do not deviate to such extent which would justify their separation.

2. Animacy. Subclasses of nouns are sometimes posited based on an animacy split, which means that animate nouns inflect for the accusative, while in case of inanimate nouns, the nominative and the accusative forms are identical. However, **Romani is not unusual in having an animacy split**. In this section, we will see this in a broader perspective and eventually eliminate the feature as one that would define separate noun classes. We will see that the split in Romani is not so clear-cut, and that the accusative form can take on other functions, too. We will also show that trying to place these problems outside the bounds of morphology does not necessarily solve them, but if we turn to constructions, it is at least easier to tackle them.

3. Case. In this section, we will first sum up Romani case marking briefly. Then we will introduce the notion of case layers as known in Indo-Aryan and show that they are not necessary and can even be misleading when applying them to Romani. We will also come to the conclusion that **the most economical is to posit two base forms for Romani nouns**, in line with the notion of stem space, and we will focus on the oblique base when looking at the weak points in the nominal paradigms.

4. Additional features. These include palatalisation and the nominative ending. Palatalisation is sometimes used to separate a certain group of nouns into two separate paradigms, but, as we will see, this is not true for Lovari on the one hand, and **palatalisation appears to be a more general tendency** in the dialects that do have this feature on the other. We will thus eliminate it as a basis for separate noun classes. The nominative ending is often claimed to be a determining factor of noun classes, but, as we will see, this is far from being so straightforward, and we will find that **the final vowel (and the lack of it) does not necessarily determine the inflectional pattern of the given noun**, so we will have to dismiss it as a definitive feature, too.

5.1.1 Gender

Gender in Romani is often said to be deducible from certain formal or semantic features of a given word, or a combination of the two, like semantic gender derivation. Although this is partly true, we will come to the conclusion that gender is, apart from the obvious cases where grammatical gender is deducible from the natural gender of the noun, is quite **arbitrary**.

5.1.1.1 Background

Among the New Indo-Aryan languages, three groups are distinguished based on the number of genders retained from Middle Indo-Aryan. One group has kept all three genders, masculine, feminine and neuter, like, for example, Gujarati in the west; the second group has completely neutralised the distinction and has no gender, like Assamese in the east, while the third group has lost the neuter but retained masculine and feminine. This is the group where Hindi as well as Romani belong. **Gender is essential in defining the two fairly distinct inflectional paradigms of the feminine and the masculine classes** and it also determines the agreement patterns of the modifiers of the noun and the obligatory selection of pronominal forms and articles.

5.1.1.2 Discussion

Elšík (2000), while giving a thorough account of the historical development of Romani nominal paradigms, notes that one of the two general criteria for all nouns is gender. In addition to this, Hutterer & Mészáros (1967) claim that one of the factors on which the classification of a noun relies heavily is the meaning of the word, or, in other words, the **natural gender**. This comes as no surprise in case of, for example, the following feminine words: *dej* ‘mother’, *žuvlí* ‘*w*oman’, *čohají* ‘witch’. The same is true for the masculine: *raj* ‘lord, master’, *juhāsí* ‘shepherd’, *bašnó* ‘cock’. Semantic gender derivation refers to feminine-masculine pairs such as the ones in Table 8.

The words meaning ‘boy’ and ‘girl’ are *čhāvo* and *čhaj*, respectively, in other dialects, with *čhaj* coming from a hypothetical form **čhāvi*. In these cases, the gender of the noun is changed through derivation, and thus, the nominal class changes as well. This is, however, no more than another form of semantic gender assignment.

masculine	gloss	feminine	gloss	origin
bāló	‘pig’	*bālí*	‘sow’	Sanskrit *bālá* ‘young’
šāvó	‘boy’	*šej*	‘girl’	Sanskrit *chāpa* ‘young one’
rom	‘man’	*romňí*	‘woman’	Sanskrit *ḍōmba* ‘man of low caste living by singing and music’
gurúv	‘bull’	*guruvňí*	‘cow’	Sanskrit *gōrūpá* ‘cow’

Table 8

Examples of semantic gender derivation

Although the natural gender of a group of words behaving in the same way morphologically can be a basis for calling that group masculine or feminine by convention, the group belonging of words without a natural gender is, rather unsurprisingly, arbitrary. It is an inherent property of every noun, and a very important one, because it is **the only one in Romani based on which we can establish true nominal classes**.

5.1.2 Animacy

A special status is attributed in Romani to the animacy of the noun, which is often referred to as an animacy split. **The animacy split presents itself in the case marking of the accusative, with animates showing an overt accusative case marker, while inanimate nouns take the unmarked nominative.** In this section, we will first describe the animacy split and the broader theory of animacy hierarchy behind it. Then we will touch upon some existing attempts at explaining the behaviour of the accusative (the **“independent oblique”** and **hyper-paradigmaticity**) and show that they only place the question outside morphology, but if we apply a constructionist approach, the problem is easier to tackle.

5.1.2.1 The animacy split

The animacy split seen in Romani is not so unusual, as, for example, in languages showing split ergativity, animacy can also have a special role among the features in a hierarchy of features, where the division in case-marking between the accusative and the ergative

appears (Silverstein 1976). Legate (2014) argues that person/animacy-based split case marking has got a morphological, rather than a syntactic source and that it is ‘synchronically encoded in the grammar’ (Legate 2014: 184). Hindi, a close relative of Romani exhibits a system of split ergativity, and, as Keine (2007) argues, it is a phenomenon of the nature of realisational morphology, and not of syntax. Silverstein’s hierarchy is reproduced here.

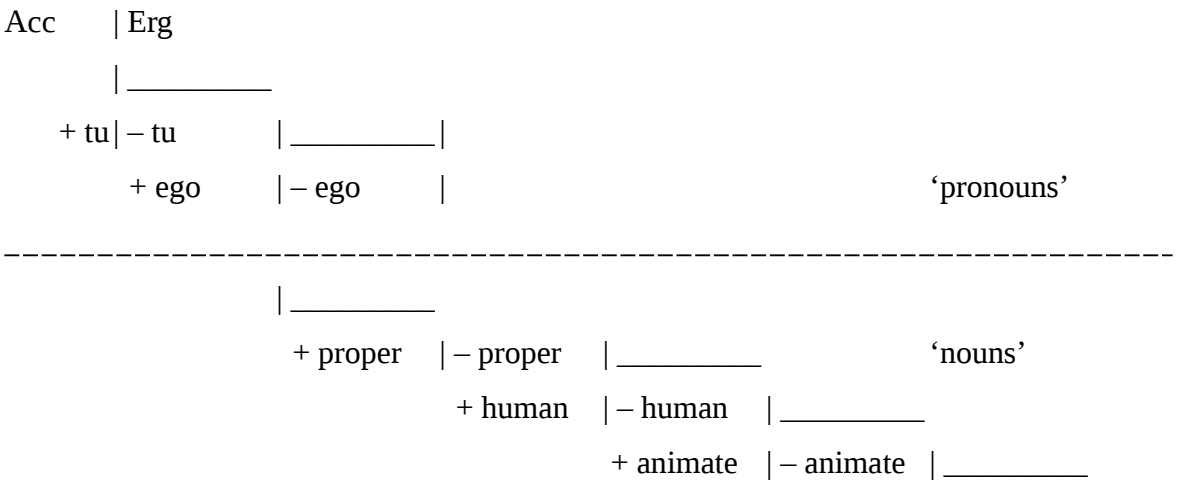

Figure 2

Silverstein’s animacy hierarchy

In expanded versions of Silverstein’s hierarchy, there are minor additions. One is that the third person pronouns and demonstratives are included, too; the other one is that kinship terms are added as a separate category between proper nouns and human nouns.

Although Romani is not an ergative language, it stands out among New Indo-Aryan languages by differentiating between third person pronouns and demonstratives (Matras 2002). From a more general point of view, this is perfectly in concordance with Silverstein’s hierarchy and its additions. Third person pronouns belong to the group of pronouns, too, and they stand higher on the scale than full nouns; therefore, if proper names have accusative marking, like in Romani (e.g. *Jã̆noš* ‘John’ ~ *Jānošés*), then third person pronouns must have it as well (*vov/voj* ‘he/she’ ~ *les/la*).

Type	Nominative	Accusative	Gloss
+ proper + human + animate	*Jă̄noš*	*Jānošés*	John
– proper + human + animate	*manúš*	*manušés*	man
– proper – human + animate	*bāló*	*bālés*	pig
– proper – human – animate	*kher*	*kher*	house

Table 9

Animacy hierarchy features in Romani

As we can see from Table 9, **the only split here is indeed between animate and inanimate nouns**. In order to describe why certain nouns inflect for the accusative and others do not, Holzinger (1993) postulated an animacy hierarchy for Romani, which is again in line with the more general hierarchy created by Silverstein (1976): from relatives through other humans and domestic animals to other animals. However, this only complicates the hierarchy further, as **we have to add features which cannot be interpreted across the whole system**.

Descriptive grammars (Hutterer & Mészáros 1967, Cech & Heinschink 1999, Boretzky 1994) all mention the difference between animate and inanimate nouns. However, they add that **in case of indefinite objects, sometimes animate nouns do not inflect for the accusative** (Cech & Heinschink 1999: 36). We can already draw the conclusion that this feature should not be considered to define separate noun classes, but we will see some more evidence in the following section.

Type		Nominative	Accusative	Gloss
+ proper + human + animate		*Jă̆noš*	*Jānošés*	John
– proper + human + animate	+ kin	*dad*	*dades*	father
	– kin	*manúš*	*manušés*	man
– proper – human + animate	+ domestic	*bāló*	*bālés*	pig
	– domestic	*māšó*	*māšés*	fish
– proper – human – animate		*kher*	*kher*	house

Table 10

Expanded animacy hierarchy in Romani

5.1.2.2 The "independent oblique" and hyper-paradigmaticity

There have been attempts to formalise this animacy hierarchy. Matras (2002) posits an "independent oblique" (which corresponds to the accusative in the examples in Table 10, for example *Jānošés*), which, besides the accusative, can also **mark other functions**, like the possessor or the benefactor of the verb 'give', as the following sentences show.

(6) *Jānošés* *si*
'John' "INDEPENDENT OBLIQUE" COPULA 3RD SINGULAR PRESENT INDICATIVE
jekh *vurdon.*
'a/one' 'car'
'John has got a car.'

(7) *De* *e* *papiňan*
‘give’ IMPERATIVE ART. DEF. OBL. geese “INDEPENDENT OBLIQUE”
te *xan*.
INFINITIVE PARTICLE ‘eat’ 3RD PERSON PLURAL PRESENT INDICATIVE
‘Feed the geese.’

Matras (2002) states that ‘the independent oblique might therefore be interpreted as consistently encoding the non-agentive referent that is high on the topicality scale’ (Matras 2002: 86). This explanation is basically identical to the reference to the animacy hierarchy, with the topicality scale being an expansion of the animacy scale (cf. e.g. Haspelmath 2004) and animates occupying a higher position on the topicality scale. He adds that ‘Romani may be said to lack a genuine accusative case altogether’ (Matras 2002: 87). If we call and define the accusative in a different way (“independent oblique”), we can get rid of an ambiguous category, but we do not necessarily explain its behaviour. Elšík (2000) also tries to ignore the question by suggesting to treat the subject/direct object split as “hyper-paradigmatic”. We should also note that there is variation here, too: both Boretzky (1994) and Elšík (2000) make mention of **purely inanimate nouns inflecting for accusative** (in East Slovak Romani: *kher* ‘house’ NOM and *kheres* ‘house’ ACC) as well as **purely animate nouns not inflecting for accusative** (in Kelderash, a Vlax dialect: *bianel e žuvli ek murš* ‘the woman bears a baby boy’), which suggests that scales (and categories) such as the animacy hierarchy might not be the best way to give account of variation. The ambiguous nature of this feature and **the confusion of animacy and inanimacy also supports the idea of dismissing it as a definitive feature** of noun classes.

In the present dissertation, I will not go into a detailed examination of the question, but I would suggest to consider this phenomenon as a weak point (see a precise definition in section 6.2), where the relationship between meaning and form becomes problematic. The primary construction contains the form noun + *-es* and the semantic content ACC. However, the same form begins to appear, pairing up with a different semantic content. As the new semantic content resembles the function of the dative, as can be seen in examples (6)-(7), we might as well call it DAT. This, in turn, results in the weakening of the form noun + *-eske/-ake/-enge/-ange*, which is the one primarily used in dative constructions. Thus, a newly emerging construction, consisting of a form taken from an existing

construction and a meaning taken from another existing construction, interferes with the existing ones, and the forms and functions become ambiguous.

5.1.3 Case

We have already seen that nominal cases exist in Romani. **In addition to the nominative and the accusative cases, there are six more: dative, genitive, locative, ablative, instrumental and vocative.** In this section, we will first give an overview of the formal characteristics of the cases. Then we will sum up why we should consider Romani case markers as suffixes rather than postpositions. Thirdly, we will show that the case layers posited for Indo-Aryan are not necessary and can even be misleading when applying them to Romani. Finally, we will come to the conclusion that there are two base forms for Romani nouns, a nominative and an oblique base.

5.1.3.1 Overview

The Romani case system is basically an exact duplicate of that of Sanskrit. In other New Indo-Aryan languages, like Hindi, the present-day system only consists of three cases: direct, oblique and vocative. While the vocative is easy to identify (the Romani vocative corresponds to the Hindi vocative), the former two are basically equivalent to the Romani nominative and accusative cases mentioned in section 5.1.2.[37] Let us see how these three cases look like exactly according to the existing descriptions.

The markers are identical for the two genders. The other five cases are marked by way of additional suffixes. Taking again the words, *tʲermó* ‘worm’ and *tʲīrí* ‘ant’, the exact forms, again in accordance with the existing descriptions of Lovari (and more generally, Vlax Romani) look like the ones in Table 12.

37 This also serves as a clue as to where the notion of the “independent oblique” originates from.

	Masculine		Feminine	
	singular	plural	singular	plural
	ʧʲermó ‘worm’		*ʧʲīrí* ‘ant’	
Nominative	*ʧʲermó*	*ʧʲermé*	*ʧʲīrí*	*ʧʲīrá*
Accusative	*ʧʲermés*	*ʧʲermén*	*ʧʲīrá*	*ʧʲīrán*
Vocative	*ʧʲérma*	*ʧʲermále*	*ʧʲîra*	*ʧʲīrále*

Table 11

Romani nominal cases which correspond to New Indo-Aryan

	Masculine		Feminine	
	singular	plural	singular	plural
	ʧʲermó ‘worm’		*ʧʲīrí* ‘ant’	
Dative	*ʧʲerméske*	*ʧʲerménge*	*ʧʲīráke*	*ʧʲīránge*
Locative	*ʧʲerméste*	*ʧʲerménde*	*ʧʲīráte*	*ʧʲīránde*
Ablative	*ʧʲerméstar*	*ʧʲerméndar*	*ʧʲīrátar*	*ʧʲīrándar*
Instrumental	*ʧʲermésa*	*ʧʲerménca*	*ʧʲīrása*	*ʧʲīránca*
Genitive	*ʧʲermésk-*	*ʧʲerméng-*	*ʧʲīrák-*	*ʧʲīráng-*

Table 12

Additional cases in Romani

We may notice that the markers of the five cases, dative, locative, ablative, instrumental and genitive[38] are all similar in that **the unique case markers are attached to a form which has so far been called the accusative**.

38 For a thorough discussion of the Romani genitive and its possible adjectival nature, see Grumet (1985) and Koptjevskaja-Tamm (2000).

case	marker	
	singular	plural
dative	*-ke*	*-ge*
locative	*-te*	*-de*
ablative	*-tar*	*-dar*
instrumental	*-sa*	*-ca*
genitive	*-k-*	*-g-*

Table 13

Romani case markers

On closer inspection, **the markers are very similar for the two numbers as well, the difference only being in voicing**. According to Matras (2002), the case suffixes whose initial vowels are /t/ or /k/ are voiced following /n/ thanks to an inherited pattern (Matras 2002: 53). Elsewhere, he refers to the same thing as voicing assimilation (Matras 2002: 88-89), which can only be interpreted in a historical phonological sense as there is no sign of voiceless stops becoming voiced in a regular manner after a sonorant in present-day Romani. It is true, however, that the consonant clusters /ng/ and /nd/ are generally more common word-internally than /nk/ and /nt/ in Romani. He also admits that 'the shape of the instrumental is less symmetrical' (Matras 2002: 89) than that of the other case markers. While the sibilant /s/ in the singular corresponds to the affricate /t͡s/ in the plural in Lovari, this is not necessarily the case in other dialects. For instance, the singular form *-ha* is also possible, like in Romungro, while the plural can take the form *-sa*. Thus, there is only thing we can state with certainty: where voiceless stops appear in the singular, it is their voiced counterparts which appear in the plural, and the sibilant does not take part in this voiceless-voiced opposition.

5.1.3.2 Case layers and base forms

Matras (2002) presents an analytical model of the Romani case system, in which the surface form of an inflected noun is assumed to consist of layers, similarly to the blocks of realisation rules in Paradigm Function Morphology (Stump 2001) and to Indo-Aryan as described by Masica (1991), referred to as Layers I, II and III.

> Romani markers of the various layers correspond in principle to those of the other New Indo-Aryan languages. This is perhaps one of the clearest pieces of evidence for a shared development of Romani and the subcontinental languages up to the New Indo-Aryan period, roughly around the tenth century AD. However, Romani case layers also show some uniqe characteristics when compared to New Indo-Aryan as a whole. The nature and position of the markers belonging to Layers I, II and III in Romani make the distinction between them more straightforward than in most New Indo-Aryan languages. (Matras 2002: 78-79)

We will see, however, that these layers are not more straightforward; in fact, they are not necessary and they are slightly confusing when discussing the morphology of Romani. **Layer I elements are composed of nominal and oblique endings**[39] which ultimately produce the nominative and accusative stems, Layer II comprises the actual case endings and, most surprisingly, a set of adpositions is called Layer III.

As can be seen in Table 14, **Layer I endings are attached directly to a nominal base**. Gender and number can be identified through the Layer I oblique ending. The question as to what exactly determines what obviously arises: is it gender that determines the oblique ending, or is it the oblique ending that determines gender? **Layer II endings are the case suffixes** mentioned in section 5.1.3.1, attached to the Layer I form of a noun. Most of the case suffixes are fixed in form (although they are subject to some variation among the dialects), showing only voice assimilation, and are added to the oblique stem.

category	form	function / meaning
lemma	*ānró*	‘egg’
nominal base	*ānr-*	stem
Layer I	*-és-*	oblique marker
	-ó	nominative ending
Layer II	*-te*	locative

Table 14

The first two case layers in Romani

39 For the sake of simplicity, I will adopt the terms “nominal” and “oblique” used in Romani linguistics in the paper, although they might as well just be dubbed short and long stems.

As we can see from the example, the terms used are ambiguous. **The use of the term "nominal" is redundant if we say that there is no other stem.** That is what we apparently see, as both the nominal and the oblique (and the vocative, for that matter) endings attach to this. Strictly speaking, the oblique "stem" is not a stem, but it is derived from the nominal base.

(8) *bakr-* + *-o* → nom. *bakró* 'sheep'
bakr- + *-es-* → obl. *bakrés-* 'sheep'
bakr- + *-a* → voc. *bakrá* 'sheep'

It would therefore be sufficient to posit one single stem which serves as the basis for all other forms of the given noun. Elšík (2000), on the other hand, proposes to differentiate between BSA (base-stem affixation) and OSA (oblique-stem affixation) languages. The former refers to languages where the cases are marked with individual suffixes; the latter means that the case suffixes are attached to an oblique stem. Romani belongs to the OSA languages, as opposed to Hungarian, for example, where case suffixes directly follow the stem, i.e. the nominative form, without mediation (nom. *bárány* 'sheep' → loc. *bárányban* 'sheep'). If we treat the nominative and the vocative independently, this could indeed be a possible analysis. Blake (2000, 2001), for example, makes reference to an oblique stem 'which serves to set the nominative off from the other cases' (Blake 2001: 42). A similar example is Lezgian (Blake 2000 based on Mel'čuk 1986), where the bare oblique stem functions as the ergative case. Elšík (2000) also mentions Daghestanian languages, where the ergative case is unmarked, similarly to Romani, where the unmarked case, as we have just seen, is the accusative, which is derived from the oblique stem by an identity process. As for the vocative, Matras (2002) notes that the vocative forms can be found 'alongside' the three layers and 'connect directly to the nominal base' (Matras 2002: 80). But even then, the nominative-oblique dichotomy remains.

From a synchronic perspective, it is more economical to say that there are two bases: the nominal base, marked by a zero morpheme: *bakr-* + *Ø*, and an oblique base, marked by a suffix: *bakr-* + *es*. This could then in turn serve as a base for the nominative

and the vocative on the one hand, and the rest of the cases on the other. Thus, we could treat all the cases in the same way, and we could avoid any reference to Layers I and II.

Although the terms Layer I and II may be justified as the characterisation of some kind of inherent property of an agglutinative language, the term Layer III is particularly misleading. Layer III in Romani derives from Indo-Aryan Layer III indeed, but while the latter is '*potentially* mediated' (Masica 1991: 234) by a Layer II element, the former has a 'preposed position' (Matras 2002: 80), and so it has no direct contact with the other layers. Let us look at examples (9)-(13).

Hindi:

(9) लड़के के साथ
laṛke *ke* *sāth*
'boy' OBL. GEN. 'with' POSTP.
'with the boy'

(10) घर के अंदर
ghar *ke* *andar*
'house' OBL. GEN. 'inside' POSTP.
'inside the house'

Romani (Lovari):

(11) *šāvésa*
šāv *es* *sa*
'boy' stem OBL. INSTR.
'with the boy'

(12) *khéreste*
kher *es* *te*
'house' stem OBL. LOC.
'in the house'

(13)*ando kher*

ande	*o*	*kher*
‘in’ PREP.	ART. DEF.	‘house’ NOM.

‘in the house’[40]

As can be seen from the examples, **there is no Layer III in Romani**. Apart from the case system presented in section 5.1.3.1, **Romani is a prepositional language**, and it uses the prepositions completely independently of the oblique base and the genitive case (which is the common “mediator” in other New Indo-Aryan languages).

It is also important to note here that, as opposed to Hindi, which is a predominantly postpositional language, Romani is a prepositional one, with the Hindi postpositions corresponding to Romani prepositions, cf. घर पर *ghar par* (H.) and *po kher* (R.) ‘on the house’. Besides or instead of the cases, several functions are or can be expressed through the use of prepositions.

angla ‘before, in front of’	*o* ‘the’	*kher* ‘house’	*anglo kher* ‘in front of the house’
andar ‘from’			*andar o kher* ‘from the house’
tela ‘under’		*kašt* ‘tree’	*telo kašt* ‘under the tree’
pala ‘behind’			*palo kašt* ‘in the tree’

Table 15

Prepositions in Romani

We have now arrived at the conclusion that **it looks economical to say that there are two base forms and a set of suffixes, some of which attach to the nominal base, while**

40 Forms such as *ande bute beršende* and *ande bute beršen* ‘in many years’ are testified in Lovari (Cech & Heinschink 1999: 18), which shows the mixing of the different ways Romani can express certain grammatical functions. Of all cases, it happens that the locative as such is actually almost never used in Hungarian Lovari and is usually replaced by the corresponding prepositional phrase, so *ando bāló* ‘in the pig’ instead of *bāléste*, *So si ande kola pohārá?* ‘What is in those glasses?’ instead of *So si kole pohāránde?*

others attach to the oblique base. This is in line with the notion of stem space as described by Bonami & Boyé (2006). There is one single lexeme equipped with a stem space with two slots.

base	form		case	form
	masculine	feminine		
nominal	*-Ø*	*-Ø*	nominative	*-Ø, -o, -i, -a, -e, -u*[41]
nominal	*-Ø*	*-Ø*	vocative	*-a/-ale*
oblique	*-es-/-en-*	*-a-/-an-*	accusative	*-Ø*
oblique	*-es-/-en-*	*-a-/-an-*	dative	*-ke/-ge*
oblique	*-es-/-en-*	*-a-/-an-*	locative	*-te/-de*
oblique	*-es-/-en-*	*-a-/-an-*	ablative	*-tar/-dar*
oblique	*-es-/-en-*	*-a-/-an-*	instrumental	*-sa/-ca*
oblique	*-es-/-en-*	*-a-/-an-*	genitive	*-k-/-g-*

Table 16

A summary of the case system of Romani

5.1.4 Additional features

In this section, we will examine two additional features, palatalisation and the nominative ending, which are sometimes considered as factors in determining the inflectional paradigm of a given noun. We will see that **the separation of feminine nouns into palatalised and not palatalised classes**, in spite of a strong palatalising tendency in general, **is irrelevant in Lovari**, and that we would have to look for other ideas to answer the question of palatalisation in other dialects. With regard to **the nominative ending**, we will show that it **is not a determining factor of noun classes**, although, as we will see in section 6.3, the word-final vowel seems to have a role in the variation in the nominal system.

41 For details about the nominative ending, see the next section.

5.1.4.1 Palatalisation

In light of what we have seen so far, it might come as a surprise that, for example in Hutterer & Mészáros (1967), **sixteen declension classes** are given for Hungarian Lovari, but Elšík (2000) also lists as many as **twelve for a reconstructed version of Early Romani**. It becomes less surprising, however, if we continue enumerating the other features determining the classes. In some varieties of Romani, a group of feminine nouns only differ in that some of them are palatalised[42] in the plural and in the oblique cases, while others are not.

(14)*žuv* 'louse' → obl. sing. *žuvá-*
suv 'needle' → obl. sing. *suvjá-*

According to Elšík (2000) and Matras (2002), this is a result of the infiltration of palatalised forms from other feminine paradigms, namely the ones with nominatives ending in a front vowel. In a number of these, the consonant preceding the front vowel is already palatalised.

(15)*pīrí* 'pot' → obl. sing. *pīrjá-*
romnʲí 'woman' → obl. sing. *romnʲá-*

We can see double neutralisation here, between the nominative forms like *žuv* and *suv* on the one hand, and between the oblique forms *suvjá-* and *pīrjá-* on the other. Elšík (2000) suggests that feminine nouns palatalised in the oblique constitute a mixed class, where the nominative form resembles that of the consonant-final feminines, like *rig* 'side' or *žuv* 'louse' in (14) as well, while the other forms are taken from feminines with a stem-final *-i*, like *pīri* 'pot' or *romnʲí* 'woman' in (15), where palatalisation is obligatory.[43]

Elšík (2000) adds that there is a tendency of depalatalisation so as to avoid mixed paradigms. However, the situation is much more straightforward than that in Lovari. **Although there are strong palatalising tendencies in Lovari, this particular form of palatalisation is completely missing from the dialect, so the oblique form of feminine**

42 This phenomenon is called "jotation" in Romani linguistics.

43 Beníšek (2012) explains the difference with the different origins of palatalised and unpalatalised nouns in Middle Indo-Aryan.

nouns is generally not palatalised.

(16)*žuv* ‘louse’ → obl. sing. *žuvá-*
suv ‘needle’ → obl. sing. *suvá-*
pīrí ‘pot’ → obl. sing. *pīrá-*
rig ‘side’ → obl. sing. *rigá-*

Therefore we can say that palatalisation is not triggered by the stem-final front vowel. **The tendency may be a more universal one**, where the alveolar stops and nasal, as well as the lateral approximant are often palatalised, irrespective of the environment. In case of feminine nouns, it is often followed by an /i/, like, for example, *romnʲí* ‘woman’, *raklʲí* ‘non-Roma girl’, *angustʲí* ‘ring’, *brādʲí* ‘bucket’, but it takes place in other environments, too, see for example *ker-* ‘do, make’ →*kerdʲom* ‘I made’. The lateral approximant is often not present at all, cf. *žuvlʲí ~ žuvjí* ‘woman’.[44]

5.1.4.2 The nominative ending

According to the descriptions and historical linguistic analyses of Romani, the nominative ending (‘the shape of the base-form suffix’ (Elšík 2000: 14), or the “Layer I markers”, as Matras (2002) puts it) plays a crucial role in the distribution of declension classes. However, we have no real reason to think that the nominative ending has a significant, let alone decisive role with regard to the nominal paradigms themselves.

The nominative ending for both genders can be any consonant, so the consonantal stems, or, in other words, the stems ending in a zero morpheme, may either be masculine or feminine. The same is true for words ending in *-i*, so there are many instances of gender neutralisation in the nominative.

(17)*phrāl* m. ‘brother’ → obl. sing. stem *phrālés-*
phen f. ‘sister’ → obl. sing. stem *pheňá-*
juhāsi m. ‘shepherd’ → obl. sing. stem *juhāsés-*
pāji m. ‘water’ → obl. sing. stem *pājés-*
patrí f. ‘leaf’ → obl. sing. stem *patrá-*

44 In Romungro, a Central Romani dialect also spoken in Hungary, the range includes the unpalatalised variety, too: *džuvlí ~ džuvlʲí ~ džuvjí* ‘woman’.

gēzeši m. ‘train’ → obl. sing. stem *gēzešés-*
butjārí m. ‘worker’ → obl. sing. stem *butjāres-*

It is already obvious that the nominative ending cannot be considered a determining factor of the declension class of a given noun. It is also clear that when the word is inflected, whatever the ending is, anything that follows the final consonant is dropped, be it a vowel, or a zero morpheme, and replaced by the oblique ending.

(18)nom. sing. *manúš* ‘man’ → obl. *manušés-/manušén-*
nom. sing. *lavutārí* ‘violinist’ →obl. *lavutārés-/lavutārén-*
nom. sing. *pheň* ‘sister’ → obl. *pheňá-/pheňán-*
nom. sing. *pīrí* ‘pot’ →obl. *pīrá-/pīrán-*

Nouns ending in *-a* are mostly feminine (e.g. *vulicá* ‘street’, *coxá* ‘skirt’, *hordőva*), but this is not universal, either: *gazda* ‘master’, obl. *gazdés-/gazdén-*. **Nouns ending in *-o*, on the other hand, appear to be exclusively masculine**: *bāló* ‘pig’, *sokro* ‘father-in-law’, *hajővo* ‘ship’.
Elšík (2000) says that the masculine nouns ending in a consonant contain a special subgroup of ‘abstract nouns, which are characterised by a specific derivational suffix’ (Elšík 2000: 14), *-ipen*.

> A unique feature of the class of abstract nouns ending in *-ipen/-iben* is the retention of a conservative masculine singular oblique form in *-ipnas/-ibnas*, rather than the expected **-ipnes/*-ibnes* (though the latter does appear as a result of a secondary development, by analogy[45] to the general masculine oblique formation, cf. Roman[46] *-ipes*). (Matras 2002: 84)

Thus, for example, the word *čačipén* ‘truth’ would have the oblique form *čačip(e)nás-* instead of the expected **čačip(e)nés-*, or, alternatively, *čačipés-*. However, *-ipes* is probably not an alternative form of **-ip(e)nes*. Rather, the loss of the final nasal of the

45 The term analogy is used here in more of an “every-day” sense, and does not have any connection to an actual analogy-based framework.
46 Roman is another Central Romani dialect, spoken in Burgenland, Austria.

suffix in Roman as well as in Lovari, and, for that matter, more generally south of the Great Divide (a bundle of isoglosses of Romani in central Europe, for more detail see section 2.5) resulted in the surface form *-ipe* (the abstract nouns thus becoming the only group with an *-e* as the nominative singular ending), and they inflect as any masculine noun: the final vowel is dropped and the oblique marker *-es-* is added.

We must make mention of the plural nominative endings, too, as it only partly corresponds to the singular ending in a regular way.[47] For nouns ending in *-i* or a consonant, the plural form ends in *-a* (irrespective of gender). In addition, with feminine nouns ending in *-i* and *-a*, we find an attractive mirror image: *-i* ↔ *-a*. Masculine nouns ending in *-o* and *-a*, however, show two different plural endings, apparently stored lexically: *-e* and *-ura*.

word	singular	plural
doktorí ‘doctor’ m.	*doktorí*	*doktorá*
manúš ‘man’ m.	*manúš*	*manušá*
bajvál ‘wind’ f.	*bajvál*	*bajvalá*
šūrí ‘knife’ f.	*šūrí*	*šūrá*
kránga ‘branch’ f.	*kránga*	*krangí*
šāvó ‘boy’	*šāvó*	*šāvé*
čókano ‘hammer’	*čókano*	*čókanura*
gázda ‘master’	*gázda*	*gázdura*

Table 17

The plural marking of nouns

5.1.5 Summary

In section 5.1, we examined some important features of Romani nouns used to classify them and we showed that the only feature which can truly determine the class of a noun is gender, which is sometimes based on certain formal or semantic features, but is generally arbitrary. We also saw that other features, such as animacy, palatalisation or the nominative

47 According to some sources, there is supposed to be a small number of masculine nouns ending in *-u*, but their existence in Lovari has not been justified without a shadow of a doubt.

ending are not straightforward enough to form the basis for separate noun classes. After looking at the case system in detail, we discarded the application of Indo-Aryan case layers, and posited two base forms for Romani nouns, a nominal base and an oblique base.

5.2 Verbal inflection

In the following sections, we will discuss the characteristics of Romani verbs, with special regard to their classification, as this plays an important role in the variation we see in the verbal system. First, we will provide a brief description of Romani verbs in general. Then we will examine some basic questions concerning the classification of Lovari verbs in the present tense and we will show that, in spite of the diachronic aspects, we must posit at least three different verb classes on a synchronic level. We will also look at the generally accepted form of the personal concord markers and suggest an alternative analysis. In addition, we will cover the problem of additional verb classes, such as the mediopassive verbs and their inflection. Moving onto the past tense, we will first present the past inflection of the consonantal class, then, as they are crucial to the better understanding of the various forms appearing in the past tense of vocalic verbs, we will discuss verb derivation and loan-verb adaptation, where we will find that the use of markers is much more restricted in Lovari than it is claimed in the literature.

5.2.1 The basic structure of Romani verbs

Without discussing the tense-aspect-modality categories in detail, let us take a quick look at **the basic structure of a Romani verb**. The morphology of the Romani verb relies on three dimensions (cf. Matras 2002: 151 and Matras 2002: 117-118): an aspectual dimension (perfective or non-perfective), a temporal dimension (remote or non-remote) and a modal dimension. The lexical roots form the core (many borrowed verbs contain loan-adaptation suffixes, as we will see in section 5.2.4.3, but in those cases the suffixed form becomes the root). **Derivational suffixes, if any, are attached to the root, creating the verb stem.** The non-perfective aspect is unmarked, the perfective aspect is marked by

a perfective marker, attached directly to the stem and creating the perfective stem.

Personal concord markers follow the stem, which may be succeeded by the markers of remoteness (pluperfect, modality) and, in the case of the present stem, of the future. Let us take an example: no derivational suffixes are attached to the root *ker-* 'do, make', so this becomes the stem. The perfective aspect is expressed by the addition of the perfective marker *-d-*: *kerd-*. This is followed by the personal concord marker: *kerdóm* (first person singular), to which we may add the remoteness marker *-as*, which renders *kerdómas*. The resulting form can express various meanings (e.g. anterior-past, request, unreal past condition) depending on the context and circumstances of usage. A descriptive look at all this in Lovari would render the matrix in Table 18 for the verb *ker-* 'do, make' in the third person singular.

tense/aspect	indicative	conditional	imperative
present	*kerél*	*kerélas*	*kér!*
past	*kerdás*	*kerdásas*	
future	*keréla*		

Table 18

Matrix of the structure of the Romani verb

The imperative is usually identical with the stem, while the conditional and the future suffixes are invariably the same, agglutinative markers: *-as* and *-a*, respectively. The form of the present conditional also expresses a progressive or imperfect aspect in the past. As variation occurs on the (present and past) stem level, and not on the level of the remoteness markers, the distribution of the stems will form the main focus of our discussion.

5.2.2 The present tense

As for the verb classes, we can say that **the classification of verbs itself is not without problems**. Depending on the point of view, verbs are put into five or two categories by the currently available literature. The traditional descriptive aspect (e.g. Hutterer & Mészáros 1967) classifies them according to the stem-final vowel (and its absence), while the

historical-diachronic approach (e.g. Matras 2002) views some of the stem-final vowels as a result of the reduction and contraction of certain derivational affixes. According to the former, there are five verb classes, whereas according to the latter, there are only two, a consonantal and a vocalic one, although at one point Matras (2002) admits that the historical processes lead to a 'reassignment of the forms from the consonantal into individual vocalic groups' (Matras 2002: 136). **Based on the newly collected data and their analysis, I will adopt a third approach by positing three fully-fledged classes.** Three additional groups of verbs will also be discussed, but partly because of their special status, partly because the data are not sufficient to evaluate them adequately, they will not be regarded as classes on a par with the other three. The three fully-fledged classes look like the following in the present indicative.

present tense indicative		consonantal class	*-a-* stem class	*-i-* stem class
number	person	*tʲin-* 'buy'	*lošá-* 'be glad'	*gindí-* 'think'
singular	1st	*tʲináv*	*lošáv*	*gindíj*
	2nd	*tʲinés*	*lošás*	*gindís*
	3rd	*tʲinél*	*lošál*	*gindíl*
plural	1st	*tʲinás*	*lošás*	*gindisarás*
	2nd	*tʲinén*	*lošán*	*gindín*
	3rd	*tʲinén*	*lošán*	*gindín*

Table 19

Verb classes in Lovari

The consonantal class is the most numerous and appears to be the most solid one, with many of its members originating from the Indo-Aryan heritage. **The *-a-* stem class has much less members** (hence, perhaps, the variation it shows in the past tense, see section 6.5.2), while **the *-i-* stem class contains most of the more recent borrowings and all of the very new borrowings**.

Based on the discussion of Burgenland Romani (Halwachs 1998), Matras (2002) draws the inference that the consonantal and the *-a-* stem classes can be deemed basic verb classes, whereas the *-i-* stem class (along with other groups of verbs to be discussed in section 5.2.2.2) is best regarded as a residual class,[48] because the forms are hard to fit into any sort of inflectional paradigm and because diachronically it came into being through derivation and the subsequent disappearance of the derivational marker (to be discussed in detail in section 5.2.4.3). However, we will see that it is not so hard to classify them after all, as the present tense of the *-i-* stem class corresponds almost exactly to the other two classes; and the past forms are more difficult to handle in case of both vocalic classes, so there is no reason not to consider it a full verb class.

5.2.2.1 Personal concord markers

The present tense personal concord markers need a bit of elaboration. According to the generally accepted analysis, the third person singular personal concord marker in the present tense of the consonantal class is connected to the stem by the linking vowel /e/, whereas in the case of the *-a-* stem class (the only "true" vocalic class according to Matras 2002) this vowel is an /a/; this renders for example the third person singular present tense form *kinél* in the case of the stem *kin-* 'buy, purchase' and *patʲál* for the verb *patʲá-* 'believe', following the assimilation of the vowel of the concordance marker. Assimilation in itself would only result in the form **patʲa-al*, therefore it will eventually be necessary to postulate that one of the /a/ vowels is deleted: *patʲa + el* > **patʲa-el* > (after the assimilation of the concord marker vowel) **patʲa-al* > (after the deletion of the concord marker vowel) *patʲál*. Alternatively, we can assume that only deletion takes place: *patʲa + el* > **patʲa-el* > *patʲál*. These two ways of derivation are valid as long as we accept that the personal concord markers are as follows (the first person forms are linked to the stem with the vowel /a/ in the case of consonantal verbs as well, which gives for instance *kináv, kinás*).

48 Based on the idea that in a historical aspect they were created subsequently; but on a synchronic level this is not relevant again.

1st sing.	2nd sing.	3rd sing.	1st pl.	2nd pl.	3rd pl.
-av	-es	-el	-as	-en	-en

Table 20

Personal concord markers according Matras (2002)

I will suggest, however, that /e/ is a default vowel[49] which is inserted when it is necessary. Thus, we find the layout in Table 21.

1st sing.	2nd sing.	3rd sing.	1st pl.	2nd pl.	3rd pl.
-av	-s	-l	-as	-n	-n

Table 21

Personal concord markers according to Baló (2008)

This renders *patja + l > patjál* in the third person singular, and the deletion of the thematic vowel or the vowel of the marker would only have to be assumed in the first persons of the *-a-* stem verbs. This analysis is also more appealing if we look at the *-i-* stem verbs: we will only have to state that the vowel of the marker is deleted.[50] The different consonant of the first person singular marker of the *-i-* stem verbs must be a result of assimilation to the thematic vowel, just like the palatalisation of the third person singular marker from /l/ to /j/. The /j/ can also be deleted completely, either accompanied by the lengthening of the thematic vowel, rendering forms such as *gindȋ* ‘I think’ or not: *rudjí* ‘I

49 The /e/ being some kind of a default vowel is also justified by the fact that it is deleted optionally or obligatorily in certain other positions. An example for the latter one is the inflexion of nouns of the *žukél* ‘dog’ type, where all other cases apart from the nominative lack the /e/ and take on the root form *žukl-*. Thus, the analysis of the personal concord markers is made easier in that we do not have to refer to assimilation in the case of vocalic verbs, but to a more general phenomenon, the role of the /e/ as a default vowel. The deletion will not take place everywhere either, only in the first persons which do behave slightly differently anyway in many cases, for example in the paradigms of consonantal verbs, too. In most cases, whether a stem belongs to the consonantal class is also made clear by the imperative, which is virtually the stem itself in the second person singular; the form is *patʲá* ‘believe (imperative)’ for the vocalic class consisting mostly of verbs with a stem-final /a/.

50 The more emphatic nature of the thematic vowel is also marked by the fact that it is never deleted at the end of the imperative form: *ker!* ‘do’ IMP., *patʲá!* ‘believe’ IMP., *gindí* ‘think’ IMP.

pray'.[51] The unusual form of the first person plural marker of the /i/ stem class (the form itself originates from a derivational marker) raises the most interesting questions, which will be analysed from several aspects.

Another solution could be to suppose that each concord marker consists of only one single consonant. This would, however, imply the unjustifiable insertion of an /a/ in the first persons of the consonantal class, as opposed to the /e/ of the other persons, which impels us to dismiss the assumption.

5.2.2.2 Additional verb classes

Besides the three verb classes discussed so far, additional thematic vowels appear in Lovari. The diachronic explanation is very similar here to the explanation mentioned in section 5.2.2 in connection with the *-i-* stem verbs. However, what will be important for us here is the special status of the first person forms, different from the rest of the paradigm.

The additional two thematic vowels are the ones that have not been mentioned yet from the basic set of Lovari vowels: /o/ and /u/. The /o/ comes from two different sources, and in one of the cases, we are probably not dealing with a separate class indeed, only the marker of a passive or middle voice.

present tense indicative	*čhinʲuv-* ~ *čhinʲó-* 'become tired'
singular	čhinʲuváv čhinʲós čhinʲól
plural	čhinʲuvás čhinʲón čhinʲón

Table 22

The present paradigm of mediopassive verbs

51 The same thing can happen with the first person singular form of the other verb classes: the /v/ can appear as any of the following stages: as a semi-vowel /w/, as part of a diphthong /aʊ/ and completely deleted accompanied by the lengthening of the vowel: *mukáv* ~ *mukáw* ~ *mukáu̯* ~ *mukā* 'I let'.

These mediopassive verbs seem to have two, alternating stems in the present. The reason behind this is said to be the contraction of the derivational marker (that is, the dropping of the consonant of the derivational suffix and the subsequent merging of the adjacent vowels), which takes place according to a hierarchy, most easily in the third person and least easily in the first person. The derived forms in the second and third persons (*čhinʲuvés, čhinʲuvél, čhinʲuvén, čhinʲuvén*) are contracted, while the derivational marker is still visible in the first persons. Matras (2002: 126, 136) propounds that the consonant of the derivational suffix is elided, and subsequently, the vowel of the concord marker (that is, the /e/) is assimilated to the vowel of the derivational suffix: *-ov-e-* > **-o-e-* > **-o-o-*. Additional deletion or fusion should be assumed to get rid of one of the two identical vowels. As for Hungarian Lovari, where the suffix takes the form *-uv*, Hutterer & Mészáros (1967) use the term "crasis" to refer to the change of the sequence *-uv-e-* to a single *-o-*. As crasis, even in its broadest sense, only involves vowels, the consonant /v/ is either deleted or becomes a vowel or a semi-vowel previously. A better derivational analysis builds upon the inventory of personal concord markers I proposed.[52] According to this analysis, the consonant of a hypothetical, underlying form *-ov-* of the suffix is deleted before consonants, that is, the second and third persons, while the vowel is raised in the first persons.

In any case, the derivational suffix and the /e/ merge into the vowel /o/, seemingly creating an extra vocalic class, but in fact, as the /o/ only appears here in the paradigm, **in case of these mediopassive verbs, there is no need to posit a separate class for them.** In addition, the striking finding in the data is that this is completely consistent: the *čhinʲuv-* variant always appears in the first persons and never appears in the other persons. This is in line with our hypothesis that /e/ is a default vowel and less stable, while /a/, the vowel of the first person markers, is only deleted when it is forced by a thematic vowel. Almost all of the verbs follow this pattern: *tordʲuv-* 'stand', *baruv-* 'grow', *ternʲuv-* 'become young', *phūruv-* 'become old', *pašʲuv-* 'lie', *najuv-* 'wash oneself', *maladʲuv-* 'meet', *dičhʲuv-* 'be seen', *sitʲuv-* 'learn', *tʲhišʲuv-* 'lose weight'. There is only one verb which does not follow suit: *paruv-* 'change'. However, it seems that the identical *-uv* sequence here is accidental, therefore it behaves just like a regular consonantal verb: *paruváv, paruvés, paruvél,*

52 I would like to thank Gyula Zsigri for suggesting the analysis.

paruvás, paruvén, paruvén.

We might want to posit that the frequency of the first person forms has something to do with the fact that they keep their original shape, but then again the frequencies of the singular and the plural forms can differ significantly. **The irregular behaviour of the first person forms is actually palpable in almost all of the verbal paradigms, including the consonantal verb class**, so it seems that the grammatical category itself also exerts a significant force, and that is why there is such a striking similarity in the differentiation.

There is another, albeit very tiny group of verbs which is similar to mediopassive verbs in that it seems to have /o/ as a thematic vowel. Attested examples of these verbs include *tʲinó-* 'shake' (from Greek *κινω* 'move'), *getó-* 'prepare' (from Romanian *găti* 'prepare'), *pahó-* 'freeze' and *mentó-* 'save' (probably from Hungarian *ment* 'save', see further discussion in section 6.5.1), and their hypothetical paradigm would look like this (their only attested form is the first person singular, and only one speaker used each).

present tense indicative	*tʲino-* 'shake'
singular	tʲinój tʲinós tʲinól
plural	tʲinosarás tʲinón tʲinón

Table 23

The present paradigm of an *-o-* stem verb

These verbs could cause a disturbance in the system and they definitely need further investigation. However, it seems that there are very few of them and they are very infrequent, so it can happen that the confused nature of the paradigm (partly similar to mediopassive verbs, partly to the *-i-* stem verbs) will eventually lead to the disappearance of these verbs.

We must note here that the historical explanation (the forms in Table 23 were created

through the loss of a derivational marker, as discussed in section 5.2.2) has to be handled with care, at least in the case of Hungarian Lovari, as the first person singular cannot be traced back to the form containing *-sar-*:

(19)1st sing. *ťinoj* ← *? ťinosarav* /##/ **ťinov*

We have no other choice but to say that the first person singular forms are based on an analogy with the *-i-* stem verbs, rather than the consonantal and the *-a-* stem verbs. There is no other pattern how to inflect a verb with a stem-final /o/, which is also closer to /i/ in height. As for the first person plural, we could presume that the aim of maintaining the paradigmatic contrast triggers a form in the first person plural which is different from the second person singular (1st plural *ťinosarás* ~ 2nd singular *ťinós*), but why is this achieved by the epenthesis of a whole sequence (a derivational marker), and why does it not happen to the *-a-* stem verbs, where the situation is the same? We find something very similar if we consider the *-i-* stem verbs, which lead us to believe that the processes must be very similar, too.

Finally, there are a handful of verbs which apparently contain /u/ as their thematic vowel.

(20)*bunu-* (< R. *bănui* 'suspect') 'regret'
sunu- (of uncertain origin) 'feel sorry'
muntu- (< R. *mântui* 'save, rescue') 'save'
rumu- (< ?Gr. *ριμαζω* 'destroy') 'go wrong'
trubu- (< R. *trebui* 'must, need') 'must, need'

Their hypothetical paradigm is as follows; they behave very similarly to the verbs like *ťinó-*.

present tense indicative	*sunu-* ‘feel pity for’
singular	sunúj sunús sunúl
plural	sunusarás sunún sunún

Table 24

The present paradigm of an *-u-* stem verb

5.2.3 The past tense

To have a better understanding of the variation of the past forms of the vocalic classes, let us have a brief look at the way the consonantal class and the mediopassive verbs inflect in the past in Lovari.

	tʲin- ‘buy’	*naš-* ‘run’	*uštʲ-* ‘get up’	*tʰov-* ‘wash’	*trad-* ‘drive’
singular	tʲindem tʲindan tʲindas	našlem našlan našlas	uštʲilem uštʲilan uštʲilas	tʰodem tʰodan tʰodas	trademtradan tradas
plural	tʲindam tʲindan tʲinde	našlam našlan našle	uštʲilam uštʲilan uštʲile	tʰodam tʰodan tʰode	tradam tradan trade

Table 25

The past paradigm of consonantal verbs

The past tense personal concord markers, which are different from the present tense personal concord markers, **attach to the past stem**, which is marked by a perfective marker. The personal concord markers, as opposed to the ones in the present tense, are universal across the verbal inflection.

1st sing.	2nd sing.	3rd sing.	1st pl.	2nd pl.	3rd pl.
-em	-an	-as	-am	-an	-e

Table 26

Past tense personal concord markers

The perfective markers are either *-d-* or *-l-*, depending on the nature of the stem-final consonant. Not all consonants appears stem-finally, probably due to accidental gaps, but generally we can say that bilabials, velars and voiceless fricatives are followed by *-l-* (with a strong palatalising tendency in Hungarian Lovari), whereas voiced alveolars are followed by *-d-*. **Palatals apparently behave slightly differently** inasmuch as is an epenthetic vowel between the stem-final consonant and the perfective marker *-l-*. However, the vowel might not be epenthetic; it could be part of a complex marker *-il-* (cf. Bubeník 2000: 214 for a diachronic analysis), and analogy at work might again be the reason for its appearance here, as this is the perfective marker of mediopassive verbs, to be discussed in this section further below.

The perfective marker *-d-* appears in case of stems ending in /d/ and /v/, too, although not as an additional element, but replacing the stem-final consonant. If /d/ behaved similarly to the other voiced stops, the stems ending in /d/ would be followed by the perfective marker *-l-*. The reason for the different behaviour might be an analogical effect again. There are numerous complex verbs ending in the verb *d-* ‘give’ (whose perfective stem is identical to the present stem, *d-*), for example: *čanga* ‘knee’ PL. + *d-* → *čangad-* ‘kneel’, *šungar* ‘salive’ + *d-* → *šungard-* ‘spit’. The influence of the very frequent verb *d-* ‘give’ can be enough in itself to be a pattern for consonantal stems ending in /d/.

The voiced labiodental fricative /v/, as we could see in Table 25, is deleted and replaced by the perfective stem marker /d/ in most cases. Further examples are *ťirav-* ‘cook’ (perfective stem *ťirad-*) or *paruv-* ‘change’ (perfective stem *parud-*), shown in (23).

(21)*parudém* *muró* *vurdón* *p'*
'change' 1st SING. PAST. IND. 1ST. SING. POSS. PRON. 'car' ACC. SING. 'on' PREP.
ek *nēvó*
'one' NUM. 'new' ADJ.
I traded my car for a new one.

The behaviour of /v/ in the past tense of verbs containing the marker -*ajv*- is slightly unusual, compared to what we have just seen about it when it is in a stem-final position: it becomes -*ajl*- in the past (*bokhajv*- 'suffer from hunger' → *bokhajl*-, *korrajv*- 'become blind' → *korrajv*-). Let us not forget, however, the complex marker -*il*-, mentioned above, and, partly more generally, partly specifically in Lovari,[53] the phonological relation between /i/ and the palatal approximant, and the frequent vocalisation or even assimilation of /v/, as seen in example (22), coming from the newly collected data: *šaravam* instead of *šaravav*.[54]

(22)*salvētésa* *šaravam* *ma*
'blanket' INSTR. SING. 'cover' 1st SING. PRES. IND. 1st SING. PERS. PRON. ACC.
ke *nā* *ma*
'because' CONJ. COP. 3RD SING. PRES. IND. NEG. 1st SING. PERS. PRON. ACC.
paplano
'duvet' NOM. SING.
'I cover myself with a blanket because I haven't got a duvet.'

Mediopassive verbs have a stem-final /v/, too (e.g. *maladʲuv*- 'meet'), so now it does not come as a surprise that their past tense is unusual. Here we find that the perfective marker is -*il*-.

53 Cf. the variation in the first person singular forms: *āldij* ~ *āldī* ~ *āldi* 'bless'.

54 Whenever there is a consonant cluster at the end of the present stem, and this consonant cluster plus the perfective marker would violate the sonority sequencing principle, the epenthetic vowel /e/ is inserted: *bistr*- 'forget' →perfective stem *bisterd*-.

past tense indicative	*malad^{j}uv-* ‘meet’
singular	*malad^{j}ilém* *malad^{j}ilán* *malad^{j}ilás*
plural	*malad^{j}ilám* *malad^{j}ilán* *malad^{j}ilé*

Table 27

The past paradigm of mediopassive verbs

The overwhelming majority of mediopassive verbs have a palatal consonant stem-finally; there is only one counterexample in Lovari, *phabuv-* ‘burn’. Verbs with a stem-final palatal tend to form their past tense with an epenthetic /i/. Therefore, **we do not have to say that this is a complex marker of any sort**. This is the perfective marker *-l-*, and we can see the combined influence of the preceding palatal necessitating the epenthesis of an /i/ and the regular deletion of any /v/ in the past tense.

The made-up derivations we will see in section 5.2.4.2 also have made-up perfective stems. As we will discuss it there, the derivation of those verbs does not conform to the semantic aspects of mediopassive verbs, and neither does their past formation: although **krajil-* and **krujil-* would be well-formed, **kucil-* would not.

5.2.4 Verb derivation and loan-verb adaptation

In order to understand in greater depth what we see in the past paradigms of the vocalic verb classes, we briefly have to look at **Lovari verb derivation markers and the adaptation of loan verbs**. Considering the analogical effects working within the domain or category of verbs, it is important to note that – due to the fact that the markers themselves end in a consonant – the derived forms are typically and generally placed in the consonantal class, no matter what their origins are.

5.2.4.1 Transitive derivational markers

There have been two main transitive derivational markers in Romani: the deverbal *-av-* and the denominal *-ar-*. The most obvious examples of the former ones (Matras 2002: 122) are listed in example (23).

(23)*darav-* 'frighten' < *dara-* 'be afraid, fear'
našav- 'drive away, expel' < *naš-* 'escape, run away'

However, there are other, less unambiguous examples. According to Matras (2002), the verb *xoxav-* 'deceive, tell a lie' derives from a non-verbal root that has been lost, while Hutterer & Mészáros (1967) claim that it has its origins in the existing verb *xox-* 'cheat, lie', which, however, does not appear in Vekerdi (2000). The verb *kirav-*[55] 'cook (tr.)' is likewise contradictory; according to Hutterer & Mészáros (1967) it derives from *kiró-* 'cook (intr.)' (and thus it is similar to the derivation *bašó-* 'make a noise' > *bašav-* 'play an instrument'), but in Matras (2002) we find that it comes from the stem *ker-* 'do, make' and it has also got another variant, *kerav-* (which thus coincides with the causative verb meaning 'have made'; Matras (2002: 123) says that the latter one is present in the Central dialects in the zone of Hungarian influence and serves as a causative marker, but according to Vekerdi (2000) it is also part of the Vlax Romani dialects, which can be the result of the effects of the Central dialects present in the area). None of the explanations is inconceivable, although both have got their own strange features. The weak point of the latter one is that it is hard to follow how the /e/ in *kerav-* turned into the /i/ of *kirav-*, although the semantic connection would not be surprising. In the case of the former one, what is problematic is the disappearance of the /o/ from the end of the stem. Then we will have to say that the suffix *-av-* truncates the stem, as the emerging vowel cluster is apparently not resolved by the insertion of a consonant. However, none of the answers is quite correct. In fact, both *kerav-* and *kirav-* 'cook' come from the same Sanskrit root, which is, however, different from the root of *ker-* 'do, make'. The verb *kiró-* (*kiruv-*) must be a mediopassive derivation containing the marker *-uv-*.

The data in (24) show that the other derivational marker, *-ar-* (which can be *-er-* or *-al-* in other Romani varieties, cf. a widely quoted example, the verb meaning 'bite', which

55 In Lovari, the form of the verb is actually *ťirav-*.

can be *dandar-* and *dindal-* in Vlax Romani or *dander-* in Vend, another dialect present in Hungary) is productive (as opposed to the marker *-av*) – the semantic content of the derived word is transparent and there are no limitations on the derivation within the given semantic field (cf. Kiefer & Ladányi 2000). Although many of these verbs are listed in dictionaries, they are not lexicalised in the sense that they acquire a genuine meaning by the addition of the marker, as shown in (24).

(24)*tató* 'warm, hot' > *tatjar-* 'heat, warm'
čókano 'hammer' > *čokanar-* 'hammer'
pají 'water' > *pajar-* 'wet v.'
čik 'mud' > *čikar-* 'muddy'
roj 'spoon' > *rojar-* 'spoon v.'

Matras (2002: 123) writes about verbs which have their origins in stems that have become obsolete, for example the phonologically interesting forms *bistar-* 'forget', *putar* 'open' and so on; these appear in Vekerdi (2000) as *bistr-* and *putr-*. The perfective stem takes the past tense (perfective) marker *-d-* after /r/, but in these cases consonant clusters are created which do not conform to the sonority sequencing principle (**-(s)trd-*) and which are resolved by the creation of forms such as *bisterd-*, *puterd-* etc. Two explanations present themselves for this phenomenon. It can be viewed from a historical viewpoint and then we can presume that it is the *-er* variant of the marker *-ar* which we can see here, while, on the other hand, if we take a synchronic approach, we might say that the consonant cluster is resolved by the insertion of the default vowel /e/. Both can be used to explain why not forms like **bistred-* and **putred-* came into being. (Apart from the forms containing the cluster *-(s)trd-* – *bistr-*, *putr-*, *inkr-* 'keep, hold' etc. –, there are no other examples for perfective forms like that, as the perfective suffix is connected directly to the stem in the case of the consonantal class, and the past tense of the vocalic verb class is formed in a different way.[56]) It is clear from the examples in (24) that the marker *-ar* is not

56 The verb *giljab-*, meaning 'sing' is interesting in that aspect and it is related to the marker *-av* in that it is one of the rare verbs which derive from nominals – in this particular case the word *gili* 'song' – with the help of the marker *-av* (it exists in the forms *giljav-* and *djilav-/djilab-* as well). Based on that, it could behave in several ways in the past tense. It could take the perfective marker *-l-*, similarly to the other bilabial, the sound /m/ (**giljabl-*) or, following the pattern of the verbs with a stem-final /v/, the /v/ (or /b/) could be deleted (**giljad-*). In the face of all that, according to Cech-Heinschink (1999) we find the

added to the nominative form but to the bare nominal stem, which does not contain any case markers.

It is important to note that although we find another marker appearing in the derivation of transitive verbs besides *-av-* and *-ar-* in the sources, the marker *-(V)sar-*, which was mentioned in section 3.1 already, the newly collected data do not confirm this (cf. *zuralo* 'strong' > **zuralosar-* 'strengthen', *phen-* 'say' > **phenosar-* 'promise'). However, it marks out a semantic difference between the verbs *vorbí-* 'talk' and *vorbisar-* 'chat', the latter expressing a diminutive or frequentative aktionsart (cf. Kiefer 2000: 292 or Schneider 2003: 15). We must add that this is the only attested instance of a semantic content attached to the marker *-(V)sar-*, which provides enough evidence for its unproductive nature.

5.2.4.2 Intransitive derivational markers

There is a marker *-av-* among the intransitive derivational markers as well, this is, however, unlike the transitive *-av-* suffix which goes back to the Sanskrit causative marker *-apaya*, cf. Matras (2002: 122), based on Masica (1991), derives from the verb *av-* 'come, become'. It is worth to note how its form is different in Lovari, perhaps as a result of differentiation from the transitive *-av-*. Its past tense form is *-ajl*, and Matras (2002) writes that it is expanded into an *-o-* stem verb in the present tense with the help of another intransitive marker, *-(j)o/u(v)-*, to be discussed in this section further below (for example *diljavo-* 'go mad' in Vlax Romani). This is inconsistent with the data found in Cech & Heinschink (1999) and Vekerdi (2000), as long as the Austrian and Hungarian varieties are concerned: we find the present tense forms *diljav-* in Austrian Lovari and *diljajv-* in Hungarian Lovari. Something similar occurs if we take a look at the verb *korav-* 'go blind', the other Vlax Romani example cited by Matras (2002: 127), which is *kor(r)ajv-* (< *kor(r)o* 'blind'), as attested by the newly collected data. **There are virtually no intransitive verbs derived using the marker *-av* in Hungarian Lovari, it is the suffix *-ajv* that appears**

---

form *djilabad-* in Austrian Lovari, which is perfectly atypical, as if the marker *-av* had doubled itself, and the form *giljabil-* in Hungarian Lovari (similarly to the verb *av-*), which seems to follow the pattern of the mediopassive verbs (see section 5.2.2.2). The seemingly unexpected appearance of past forms in *-il-* in certain cases may be a result of an analogy with verbs expressing mental aspects, but it is worth noting that verbs with a stem-final palatal sound, which belong to the consonantal or the vocalic class but where no mental aspects are involved also take the *-il-* suffix: *tordjo-* 'stand', perfective *tordjil-*, *uštj-* 'stand up', perfective *uštjil-*, *ašo-* 'stay', perfective *ašil-*.

everywhere, including the present tense, as confirmed by the data shown in (25).

(25)*xōjí* ‘anger’ > *xojajv-* ‘be angry’
bokh ‘hunger’ > *bokhajv-* ‘suffer from hunger’
zūr ‘strength’ > *zurajv-* ‘become strong’

Similarly to the markers in section 5.2.4.1, this marker is not linked to the nominative form but to the bare nominal stem, without any case markers. **The grounds for the spread of the past tense form onto the present tense (the perfective stems are indeed *xojajl-*, *bokhajl-* and *zurajl-*) could have been the desire to maintain the transitive-intransitive contrast.** In certain varieties, the simultaneous presence of variants containing the marker *-ajv-* and forms without it could result in a differentiation between the meanings, and in such cases the latter one will be non-inchoative, as shown in (26).

(26)*loša-* ‘be glad’ < *loš* ‘joy’ > *lošajv-* ‘become glad’
langa- ‘limp’ < *lango* ‘lame’ > *langajv-* ‘become lame’
dukha- ‘hurt’ < *dukh* ‘ache, pain’ > *dukhajv-* ‘become painful’

The other intransitive marker, the mediopassive marker, originally bore the form *-(j)o/u(v)-* and presumably derives from the verb *ov-* ‘become’, which still exists in the Romungro dialects. The marker appears in Lovari in the form *-uv-*, and virtually the /j/ is kept, too, by the palatalisation of the stem-final consonant preceding it (cf. Hutterer & Mészáros 1967). As mentioned in section 5.2.2.2, the marker, in effect, only appears in the first persons. Although semi-conscious language planning may employ it throughout the whole present tense paradigm, like in the examples in Table 28[57], this is not attested in any source of data at all (but see *paruv-* ‘change’ in section 5.2.3). Its productivity in general is a question yet to be answered. The verbs in the table clearly show a misinterpretation and therefore an ad hoc use of the mediopassive marker to create made-up derivations.

57 Information provided by Szilvia Lakatos, lecturer at the Department of Romani Studies at the University of Pécs.

present tense indicative	*kraj* ‘king’ > *krajuv-* ‘rule’	*krujal* ‘around’ > *krujuv-* ‘go round’	*kucuv-* ‘whet’[58]
singular	krajuvav krajuves krajuvel	krujuvav krujuves krujuvel	kucuvav kucuves kucuvel
plural	krajuvas krajuven krajuven	krujuvas krujuven krujuven	kucuvas kucuven kucuven

Table 28

Made-up derivations in Lovari

According to some sources, there is a third marker, too, which appears in intransitive derivation, namely *-sajv-/-(V)s-av-*, with the past form *-sajl-* (cf. *-ajv-* and *-ajl-* above), e.g. *kolo* ‘soft’ > *kolosajv-* ‘become soft’, *lolo* ‘red’ > *lolosajv-* ‘turn red’, *lungo* ‘long’ > *lungosajv-* ‘become longer’; but this was not confirmed by the newly collected data.

Summing up the derivational markers now, we find the relatively symmetric picture shown in Table 29. In spite of the fact that, on the one hand, Matras (2002) says that the markers in the grey cells are loan-verb adaptation markers (see section 5.2.4.3), and that, on the other hand, I previously argued (cf. Baló 2011 and Baló 2012) that they take part in internal derivation, they seem to have been recycled for a different function.

transitive markers			intransitive markers		
-av	-ar	-sar	-ajv	-uv	-sajv

Table 29

Matrix of verb derivation in Lovari

58 In Vekerdi (2000), we actually find *-i-* stem verbs with the same meanings: *kruji-* ‘go round’ (< Serb. *krug* ‘round’) and *kuci-* ‘whet’ (< Rom. *ascuți* ‘sharpen’), which naturally conforms to those said about the made-up stems.

5.2.4.3 Loan-verb adaptation

Many (Bakker 1997, Matras 2002, Boretzky 1994, Choli-Daróczi & Feyér 1988, Cech & Heinschink 1999, among others) have claimed that borrowed lexical items in Romani are marked out by special derivational markers, and thus, the inherited and the borrowed parts of the lexicon form two, grammatically different layers. As we have seen already, this is clearly not the case, and what belongs to the "borrowed" part is highly controversial and extremely hard to determine anyway. Although it would be interesting to know, for example, whether the words and grammatical markers which became part of the language during the lengthy sojourn of the Roma in Byzantium, before they scattered in Europe and the diversification of dialects began, belong to the core lexicon or cannot be deemed equally influential or basic as the Indo-Aryan vocabulary. Psycholinguistic research has yet to establish how these layers are regarded by native speakers and how they are actually stored in the mind, because it may influence the analogical processes at work in the verbal system. **For the purposes of the present study, and due to a lack of sufficient evidence, we will disregard the possible differences resulting from the temporal aspects of the lexicon and consider all forms as carrying equal weight.**

The origins of the loan-verb adaptation markers go back to the Greek inflection endings; for example in Vlax Romani it is the Greek aorist forms (*-is-/-as-/-os-*) that appear (Bakker 1997, Boretzky & Igla 1991). However, they do not insert the new, borrowed verbs just by themselves: they are linked to the derivational markers *-ar* and *-av* discussed in section 5.2.4.1, depending on whether it is a transitive or an intransitive verb. This is where the markers *-sar* and *-sajv* come from, and this is how they can be broken down on a historical basis into a "carrier" derivational marker (*-ar* and, as we could see in section 5.2.4.2, *-ajv* instead of *-av*) and the suffix *-(V)s-* which would serve to mark the fact that the verb is borrowed. As the markers themselves end in a consonant, too, all the loan verbs formed by their addition are inserted into the consonantal class which has got the highest type and token frequency. This is how they are supposed to look like – but they just do not exist in this form in Hungarian Lovari.

(27)Hun. *ás* > **ašisar-* 'dig'
Rom. *gîndi* > **gindisar-* 'think'
Hun. *indul* > **indulisar-* 'leave'

Rom. *scrie* > **iskirisar-* ‘write’

Rom. *ajuta* > **žutisar-* ‘help’

Gr. *χανω* (aor. *χασα*) > **xasajv-* ‘disappear’

Hun. *kezdődik* > **kezdēdisajv-* ‘begin’

Rom. *gîndi* > **gindisajv-* ‘think’

Rom. *scăpa* > **skepisajv-* ‘escape’[59]

Instead, we find the forms in Table 30.

**iskirisar-*	*iskirí-*
**indulisar-*	*indulí-*
**žutisar-*	*žutí-*
**kezdēdisajv-*	*kezdēdí-*
**gindisajv-*	*gindi-*

Table 30

Loan verbs in Lovari

Here, we must also make mention of a marker we will see in section 6.5.2.1, *-in-*, which goes back to the Greek present tense inflection markers (Bakker 1997: 128). In Cech & Heinschink (1999), we saw that the marker is used to mark out loan verbs. According to the newly collected data, in Hungarian Lovari, this marker is not used in this function either, although Hutterer & Mészáros (1967), quoting, among others, the following examples, list it as a loan-verb adaptation marker: *bokszol* > *boksolin-* ‘box’, *szív* > *sivin-* ‘suck’, *arat* > *aratin-* ‘reap’.

We can say that loan-verb adaptation as described by our sources and the existing descriptions does not exist in Hungarian Lovari. Loan verbs are not specifically marked, just like the way we saw it in the case of nouns. We have to say that **loan-verb adaptation follows certain patterns, and a strong pattern is represented by the *-i-* stem verbs**, but in other dialects and varieties, the consonantal class exerts its

59 In the sources, for example Vekerdi (2000), we also find forms such as *getosar-/getosajv-* ‘prepare’ (~ *gető*, see the discussion in section 5.2.2.2).

influence, too, by employing certain markers which place the new verbs in the class of the highest type and token frequency.[60]

5.2.5 Summary

In this section, we discussed the characteristics of Romani verbs. We posited three verb classes in Lovari, the consonantal class and two vocalic classes, one with a stem-final /a/ and another one with a stem-final /i/. This classification is crucial to the understanding of the variation in the verbal system. We also looked at the conjugation of mediopassive verbs and examined the question of other possible stem-final vowels and verb classes. While describing the present tense of verbs, we also suggested an alternative analysis of the present tense personal concord markers.

In order to understand the variation described in Chapter 6, we looked at the past tense of consonantal verbs, verb derivation and loan-verb adaptation, where we found that the use of loan-verb adaptation markers is much more restricted in Lovari than it is claimed in the literature.

60 A somewhat similar phenomenon can be seen in Daco-Romanian (cf. Costanzo 2008), where loan verbs of Balkan origin mostly fall into the [+sc] subclass of the 4th conjugation, but that is not always the case; it may happen that they fall into the [-sc] subclass or a different conjugation altogether. This variation continues into the contemporary language, as shown by the example of the English verb *blog*, which can be *bloguiesc* but also *bloghez*. Costanzo (2008) adds – and this is true for Lovari, too – that different patterns are employed and that variation is a result of analogical change.

6 Variation in Lovari morphology

In this chapter, we will first make an attempt at defining the notion of a weak point and provide a preliminary outline of the three weak points in Lovari morphology under discussion. Then we will describe the three weak points in detail and introduce the possible analogical forces behind the variation seen at these points.

6.1 The notion of a weak point

In order to clarify what a weak point is,[61] we will use the idea that **the regularities on a particular level of linguistic description can be expressed in terms of schemata** (Booij 2010, following the notion of schema, as described by Rumelhart 1980). Although, closely related, schemata represent a more general notion than constructions. While the latter denote a pairing of form and meaning (Goldberg 1995, Jackendoff 2008), the former, in case of morphological schemata, contains **phonological, syntactic and semantic information**.[62] For example, the schema for deverbal *-er* in English is as follows, where the symbol ↔ stands for correspondence (Booij 2010: 8).

61 A weak point is in fundamentally similar to an unstable point, as defined by Rebrus & Törkenczy (2011). They define an unstable point in paradigms as 'those points in the paradigm where more than one conflicting analogical requirement applies with approximately equal strength' (Rebrus & Törkenczy 2011: 139). Although the present paper will mainly deal with formal connections, they add that a functional relationship can also serve as an analogical connection.

62 Rebrus & Törkenczy (2005) do something similar when they underspecify the input in the framework of Optimality Theory by defining its morpho-syntactic characteristics only and rely on output-output constraints to determine the outcome of two cases of lexical allomorphy in the Hungarian verbal paradigm. The two cases are Definiteness Neutralisation and Anti-Harmony, and the constraints they use require paradigmatic uniformity on the one hand and paradigmatic contrast on the other. We may say that, in some way, the underspecified inputs correspond to the semantic and the morpho-syntactic component of the schemata, while the correspondences between the components of a schema or between components of different schemata are similar to the ranking of the constraints.

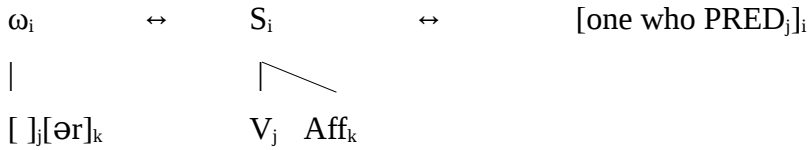

Figure 3

Schema for deverbal *-er* in English

The three kinds of linguistic information included here are the phonological form ω, the syntactic information S (that it is a deverbal affix), and the semantic information. Similarly, the schema for the Hungarian plural suffix *-k* would be the one shown in Figure 4.

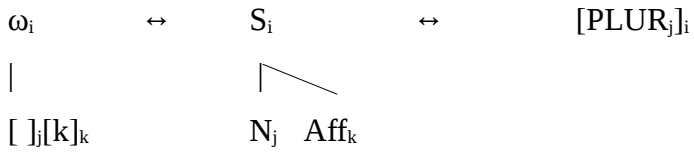

Figure 4

Schema for the Hungarian plural suffix *-k*

Instead of this representation, based on the idea of Booij (2010), I suggest **a circular representation** of the schema, as sketched in Figure 5, where every kind of information is connected to the other two through correspondences, marked by arrows in both directions, as there is a relationship between the semantic and the phonological information as well.

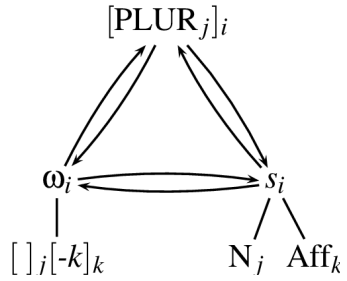

Figure 5

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 overall strength of the schema, which would in turn 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 Lovari, where the semantic component may pair up with a different set of phonological form and syntactic position, resulting in variation. Thus, a weak point in morphology is a schema where at least one of the correspondences is not mutually unambiguous.

We can draw up the following, combined schema, shown in Figure 6, consisting of two schemata, for the locative case in Lovari. The upper section of the schema describes the agglutinative case marking: it contains the phonological form, for instance, taking an example we have encountered already, *kheréste*; the morpho-syntactic information, which says that the case affix is attached to the oblique base of the noun; and the semantic component, which is the locative function in this case.

Sticking with the same example, we might recall that there is an alternative way of expressing the locative, by means of a preposition: *andó kher* (the form is in fact made up of the preposition *andé* and the definite article *o*, but that is irrelevant here; the article immediately precedes the noun in every case, so the other form of the locative with case marking and including the definite article is *e kheréste*). The lower section of the schema shows the prepositional locative.

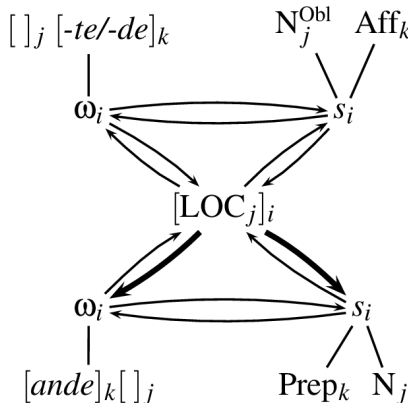

Figure 6

Schema for the locative case in Lovari

The thick arrows in this schema mean that the correspondences in that direction prevail in the expression of the locative case, so the prepositional form is more typical than the agglutinative one. 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 the way of marking commonly called “regular” (the addition of the suffix *-ed*). If all English verbs inflected that way, there would only be one 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* etc.). The existence of these verbs means that the correspondence between the past tense function and the marker *-ed* is not unambiguous, and neither is the correspondence between the past tense function and the morpho-syntactic property of affixation for the past tense. Several other morpho-syntactic ways 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*).

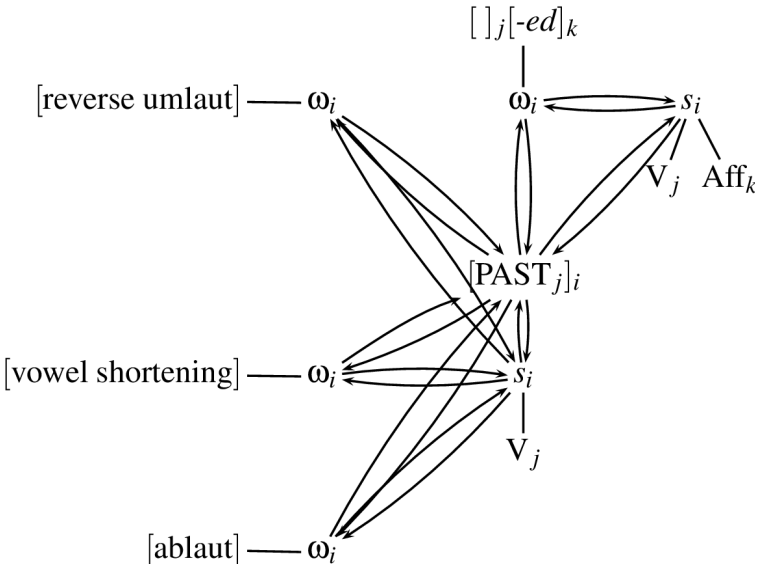

Figure 7

Schema for the English past tense

With so many schemata 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 probably depends on other factors, such as frequency, the extent of the embedded nature of the forms etc. However, if variation emerges, 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 morpho-syntactic influence on the forms that begin to vary.

6.2 An overview of the weak points in Lovari

There are **three weak points** in Lovari inflection where variation occurs[63] and where the surface forms (surface similarities and differences; in general, cf. e.g. Kálmán, Rebrus & Törkenczy 2012) and analogical effects might play a role in producing and maintaining this variation. Let us have a look at them one by one.

1. The first weak point we will look at is **the masculine oblique base**. As discussed in section 5.1.3, the 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 schema does not exclusively prevail within the masculine nouns. It is weakened by the existence of another phonological form, containing *-os-* in the singular and *-on-* in the plural, so, for example, the oblique forms of the word *fő́ro* 'town' are *fōrós-* and *fōrón-*, respectively.
2. The second weak point can be found in the feminine class of nouns. The oblique marker in the singular is invariably *-a-*: *šej* 'girl' ~ *šejá-*, *žuv* 'louse' ~ *žuvá-*. As seen in sections 5.1.1 and 5.1.3, the feminine plural oblique marker is *-an-*, for example the plural oblique base of *šej* 'girl' is *šeján-*. However, there is another phonological form of **the feminine plural oblique marker**, *-en-*, see for example *žuv* 'louse', whose plural oblique base is *žuvén-*.
3. The third weak point can be found in the verbal system of Lovari. The past tense of consonantal verbs shows a pretty regular nature, as we have seen in section 5.2.3.

63 We must note that the present paper does not deal with the possible diachronic processes that could have led to this variation and are emphasised heavily in the literature on Romani linguistics.

However, there are verbs which are different from them, as discussed in section 5.2.2, in that their stem ends in a vowel, and not a consonant, and there is no existing, straightforward pattern for these verbs. With an unambiguous pattern missing, we will see that the past tense of consonantal verbs will only be one of the patterns used for **the past form of vocalic verbs**.

6.3 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 go over six 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. They are the following.

1. The position of stress. At first glance, it seems that there is at least some sort of correlation between the variation of the oblique forms and the fact that Lovari lacks a straightforward stress pattern. Stress itself seems 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 pairs up with the unpredictability of oblique forms.
2. The 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 three (words with four or more syllables are rare, although we will see some examples in section 6.3.1). The variation is especially ostensible on trisyllabic words with a stem-final /o/, while disyllabic words never vary.
3. The plural form. There are two possible nominative plural endings for masculine nouns. It seems that the plural ending can provide us with some clue as to the distribution of the oblique ending, but it can be predicted only partially. In addition, we must also note that the nominative plural form cannot be predicted unambiguously from the nominative singular.
4. The masculine adjectival ending *-ano*. There is a set of denominal adjectives whose

ending is *-ano* in the nominative and *-ane* in the oblique. These adjective are interesting because their nominative ending is identical to certain nouns which show a high degree of variation, but also more generally, due to the fact that their oblique form ends in /e/, while their nominative form ends in /o/, which is similar to what we find in one of the patterns for the masculine oblique base, where nouns ending in /o/ take the oblique forms *-es-/-en-*.
5. 2nd person singular verbal endings. The 2nd person singular present indicative ending of many verbs (those with a stem-final consonant) is *-es*, while that of the mediopassive verbs is *-os*. We will examine whether there is any correlation between the proportion of the type frequency of the consonantal verbs and the mediopassive verbs and the proportion of the masculine nouns with the oblique ending *-es-/-en-* and the oblique ending *-os-/-on-*.
6. The adverbial ending *-es*. Although the data here are particularly scarce, as a last possibility, we will briefly discuss whether the adverbs mostly derived from adjectives and ending in *-es* can influence the choice of the singular oblique ending in the masculine in favour of *-es-*, as opposed to *-os-*.

6.3.1 Description of the phenomenon

In this section, we will introduce **the variation in the masculine oblique base** and we will also see that this variation **is closely linked to the masculine nouns which have a stem-final /o/**.

So far, we have seen one suffix for the oblique singular base and one for the oblique plural base, which seems to attach to all masculine nouns: *-es-* and *-en-*, respectively. But this is not true for all masculine nouns. There are ones which, without any apparent phonological or morpho-phonological reason, take a different oblique marker: *-os-* in the singular and *-on-* in the plural. This means that **there are two "sets" of oblique markers in the masculine**: *-es-/-en-* and *-os-/-on-*.

(28)*hī̆ro* 'a piece of news' → obl. *hīrós-/hīrón-*
fŏ́ro 'town' → obl. *fōrós-/fōrón-*

Masculine nouns can be divided into three groups according to the oblique form: in the first group, only the oblique in *-es-/-en-* is used, in the second group, only the the oblique in *-os-/-on-*, and there is a third group where the two possible forms vary. The two competing patterns can be seen here next to each other throughout the whole paradigm in Table 31.

masculine	*bakró* 'sheep'		*sókro* 'father-in-law'	
	singular	plural	singular	plural
N	bakró	bakré	sókro	sokrurá
A	bakrés	bakrén	sokrós	sokrón
D	bakréske	bakrénge	sokróske	sokrónge
L	bakréste	bakrénde	sokróste	sokrónde
Abl	bakréstar	bakréndar	sokróstar	sokróndar
I	bakrésa	bakrénca	sokrósa	sokrónca
G	bakrésk-	bakréng-	sokrósk-	sokróng-
V	bákra	bakrále	sókra	sokrále

Table 31

The two masculine paradigms

We can draw up the following schema, shown in Figure 8, for the masculine oblique base, where N is a masculine noun. It contains the oblique marker *-es-/-en-* as the phonological form on the one hand, and the oblique marker *-os-/-on-* on the other.

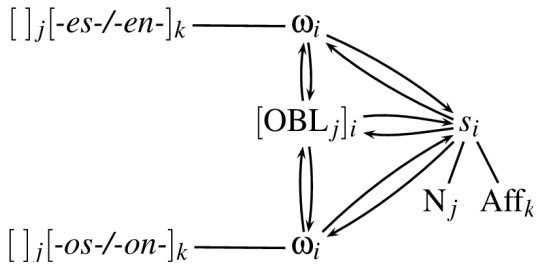

Figure 8

Schema for the masculine oblique base

In this combination of two separate schemata, 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 is weakened by the presence of the other schema, where the same semantic content corresponds to a different phonological form, $\omega_i[\ ]_j[os/on]_k$, and this is also true the other way round: **the correspondences between each phonological form and the semantic content OBL_j are weakened by each other**.

To illustrate this, Tables 32-34 sum up the masculine nouns that can be found in Vekerdi (1985). Only words with at least one attested oblique form are included. The first list contains 67 lexical items with the oblique form *-es-/-en-*, the second one contains 17 masculine nouns with the oblique form *-os-/-on-*, while the third table contains only 1 item, the one where there is variation. In the tables, the words are grouped together in the order of the number of syllables (nouns with one syllable only appear among the ones with the oblique form *-es-/-en-*, while nouns with four syllables only appear among the ones with the oblique form *-os-/-on-*). Within the groups, the words are listed according to the end of the stem: whether there is a consonant, an /i/ or an /o/.

noun	attested oblique forms
one syllable	
bal ‘hair’	*balénca*
beng ‘devil’	*bengéske*
berš ‘year’	*beršénca/beršénge/beršéngo/beršéngi/beršésa*
dad ‘father’	*dadéske/dadés*
del ‘god’	*devlésа/devléstar/devlés*
djes ‘day’	*djeséstar*
drom ‘road’	*dromésk*e
gad ‘shirt, clothes’	*gādénde*
gav ‘village’	*gavésko/gavéste/gavénge*
grast ‘horse’	*grastés/grastéski/grasténca/grasténgo/grastén/grasténge*
kham ‘sun’	*khaméski*
kraj ‘king’	*krajéski/krajéska/krajéske/krajésko/krajéstar/krajéngo/krajénca/krajén*
nakh ‘nose’	*nakhéski*

phāk ‘wing’	*phākéngo*
phral ‘brother’	*phralés/phralénca*
raj ‘lord’	*rajénge*
rat ‘blood’	*ratésa*
rom ‘man’	*romés/roméske/roméste/roménca/roménge*
rup ‘silver’	*rupésa*
ruv ‘wolf’	*ruvéske*
sap ‘snake’	*sapén*
šon ‘month’	*šonéngo*
them ‘country’	*theménge/theméngo*
vast ‘hand’	*vasténde/vasténgi/vasténdar/vasténca*
veš ‘forest’	*vešénde*
two syllables	
ākhór ‘walnut’	*ākhorén*
ambról ‘pear’	*ambrolén*
dudúm ‘pumpkin’	*dudumésk*e
manuš ‘man’	*manušéske/manušénca/manušén*
rašáj ‘priest’	*rašajésko*
šošoj ‘rabbit’	*šošojés/šošojéngi/šošojénge/šošojén*
vurdón ‘car, carriage’	*vurdonénca*
žukél ‘dog’	*žuklésk*e/*žuklénca/žuklén*
čə̄ri ‘sky’	*čērésk*e
kə̄rtji ‘card’	*kārtjénca*
pājí ‘water’	*pājéste/pājésa*
ānró ‘egg’	*ānrés*
bakró ‘sheep’	*bakrés/bakrén*
bāló ‘pig’	*bālésk*e/*bālésa/bālés*
bašnó ‘cock’	*bašnésko*
dumó ‘back’	*dumésa*
gāžó ‘non-Romani man’	*gāžésk*e
gōnó ‘sack’	*gōnés*
xānró ‘sword’	*xānrésa*
lōvé ‘money’ pl.	*lōvénge*
mānró ‘bread’	*mānréski*

māšó ‘fish’	*māšéske/māšés*
punró ‘foot, leg’	*punréngo*
rakló ‘boy’	*raklés/raklésk*i/*rakléske/raklénge*
šāvó ‘boy’	*šāvés/šāvéstar/šāvéske/šāvésko*
šēró ‘head’	*šēréste/šērésa/šērésko/šēréngo*
three syllables	
bēréši ‘farmhand’	*bērešéske*
bójtāri ‘shepherd boy’	*bojtārénge*
čapóši ‘bartender’	*čapošés*
farkáši ‘wolf’	*farkašéski*
fillḗri ‘penny’	*fillērésko*
xanrāló ‘policeman’	*xanrālénge*
juhā̆si ‘shepherd’	*juhāsés/juhāséske*
kirā̆ji ‘king’	*kirājés/kirājéstar/kirājéske/kirājésko*
kóčiši/kočíši ‘coachman’	*kočiséske/kočisés*
murmúnci ‘cemetery’	*murmuncéngi/murmuncéngo*
šārkān^{j}i ‘dragon’	*šārkān^{j}és/šārkān^{j}éske/šārkān^{j}éstar*
téngeri ‘sea’	*tengéreske*
padlṓvo ‘floor’	*padlōvésko*
raklōró ‘little boy’	*raklōrén*
šāvoró ‘little boy’	*šāvorés/šāvoréngo/šāvorén/šāvoréngi/šāvorénca/ šāvorénge*
žamutró ‘son-in-law’	*žamutrés-*

Table 32

Masculine nouns with the oblique form *-es-/-en-* from Vekerdi (1985)

noun	attested oblique forms
two syllables	
bógo ‘nag’	*bogós*
čā̆so ‘hour, watch’	*čāsóngo*
djā̆so ‘mourning’	*djāsóste*
grṓfo ‘count’	*grōfóske*
pápo ‘grandfather’	*papósko*

pa̋to ‘bed’	*pātósko*
pérco ‘minute’	*percónde*
práho ‘dust’	*prahóske*
púјo ‘chicken’	*pujón*
sóbro ‘statue’	*sobrón/sobrónde*
three syllables	
āláto ‘animal’	*ālatós/ālatón/ālatónge*
bárāto ‘friend’	*barātóske*
čala̋do ‘family’	*čalādóske*
hércego ‘prince’	*hercegón*
rablő́vo ‘robber’	*rablōvóske*
unóko ‘grandchild’	*unokósko*
four syllables	
sörnjetégo ‘monster’	*sörnjetegóski*

Table 33

Masculine nouns with the oblique form *-os-/-on-* from Vekerdi (1985)

vité́zi ‘brave warrior’	*vitēzón/vitēzós/vítēzés/vítēzéske*

Table 34

Masculine noun where there is variation from Vekerdi (1985)

The masculine nouns we have from the newly collected data are listed in Tables 35-37.[64] Again, only items which have at least one attested oblique form were taken into consideration. The tables contain 28 masculine nouns whose oblique form is *-es-/-en-*, 23 masculine nouns whose oblique form is *-os-/-on-*, and, in addition, there are 8 lexical items whose oblique forms vary. On the one hand, we can see that the proportion of the two oblique forms has changed and become more balanced. On the other hand, there are more stems which vary.

64 Due to the complete lack of the locative case mentioned before, we lost a lot of potential data.

noun	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’	*kherésk*e/*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	
gẽzeši ‘train’	*gēzešésa*
koldúši ‘beggar’	*koldušéstar/koldušés/koldušén/koldušénca*
kopə̃či ‘tree trunk’	*kopāčéske*
pohə̃ri ‘glass’	*pohārénca*
four or more syllables	
kočmāróši ‘bartender’	*kočmārošénca*
kirčimārúši ‘bartender’	*kirčimārušésa/kirčimārušéstar*

Table 35

Masculine nouns with the oblique form *-es-/-en-* from the newly collected data

noun	attested oblique forms
two syllables	
ā*tko* ‘curse’	*ātkónca*
búso ‘bus’	*busósa*
čā̆so ‘hour, watch’	*čāsóngo*
fő̆ro ‘town’	*fōróske*
gíndo ‘problem’	*gindóstar/gindónca*
hī̆ró ‘a piece of news’	*hīróstar*
nā̆so ‘child’s father-in-law’	*nāsósko*
nī̆po ‘relatives’	*nīpósa/nīpós*
*púj*o ‘chicken’	*pujón*
ríto ‘field’	*ritóske*
trájo ‘life’	*trajósko*
three syllables	
āláto ‘animal’	*ālatón/ālatós*
bārő̆vo ‘baron’	*bārōvóske*
čalā̆do ‘family’	*čalādós/čalādósa/čalādón*
falató ‘a little bit of food’	*falatóske/falatón*
xāmásko ‘food’	*xāmaskós*
jōsā̆go ‘livestock’	*jōsāgós*
laptópo ‘laptop’	*laptopósa*
sómsē̆do/sómsī̆do ‘neighbour’	*somsēdósko/somsēdóski/somsīdós/somsēdós/somsēdóske*
vonáto ‘train’	*vonatósa*
four or more syllables	
ternimā̆ta ‘the young ones’	*ternimātós/ternimātóske/ternimātónca/ ternimātónge*
šegīččē̆go ‘help’	*šegīččēgóske/šegīččēgós*
sā̆mītōgē̆po ‘computer’	*sāmītōgēpósa*

Table 36

Masculine nouns with the oblique form *-os-/-on-* from the newly collected data

noun	attested oblique forms
two syllables	
sókro ‘father-in-law’	*sokróske/sokrónge/sokrénge*
three syllables	
čókano ‘hammer’	*čokanésko/čokanósko*
dúhano ‘tobacco’	*duhanés/duhanéski/duhanós/ duhanóski*
kirãji ‘king’	*kirājéske/kirājénge/kirājén/kirājón*
mobílo ‘mobile phone’	*mobilésa/mobilósa*
pokrőco ‘blanket’	*pokrōcésa/pokrōcósa*
four syllables	
kirčimãri ‘bartender’	*kirčimārésa/kirčimārósa/kirčimāréstar/kirčimāróstar/kirčimārénca*
telefóni/telefóno ‘telephone’	*telefonésa/telefonósa*

Table 37

Masculine nouns where there is variation from the newly collected data

There is an overlap between the two lists, so all in all, we have 82 lexical items with the oblique form *-es-/-en-* and 36 items with the oblique form *-os-/-on-*. The total number of stems whose oblique forms vary is 9. In some cases, the variation is slight, with one or the other more dominant, but there are cases, like *dúhano*, where we find that the amounts of the two different oblique occurrences are equal.[65] The overall proportion of the frequency of the stems with the oblique forms *-es-/-en-*, *-os-/-on-* and the stems where the forms vary looks like this.

65 More evidence for the variation comes from Cech *et al.* (1999), where we find a further example: the oblique form of the word *kókalo* ‘bone’ appears as both *kokalós-* and *kokalés-*.

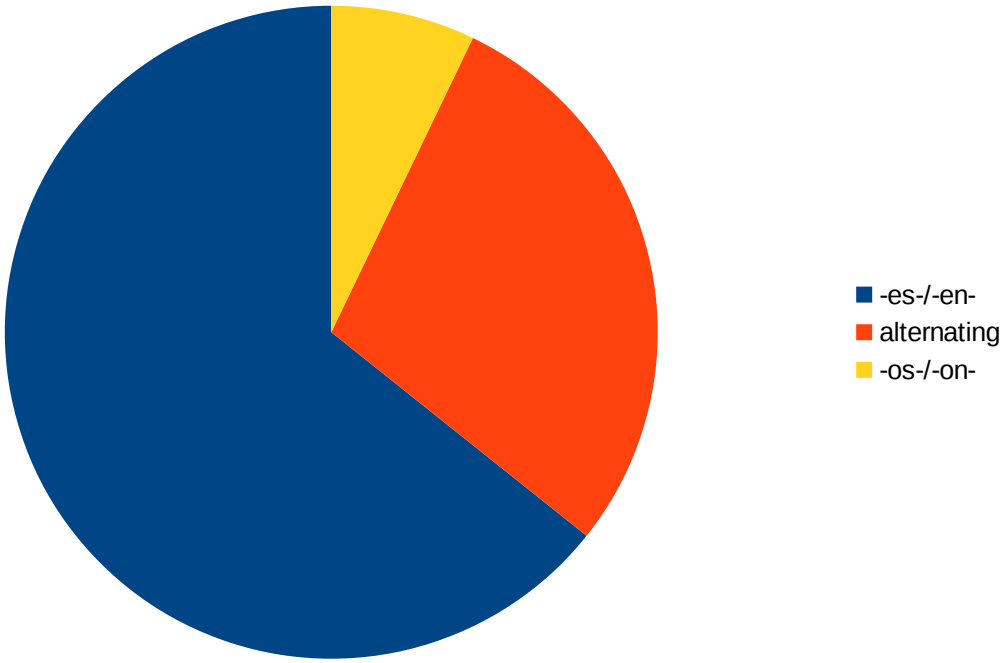

Figure 9

Proportion of the frequency of the stems with the oblique forms *-es-/-en-*, *-os-/-on-* and the varying stems

Variation seems to appear more often among words where the final vowel of the nominative singular form is /o/: *čokáno* ‘hammer’, *dúhano* ‘tobacco’, *mobílo* ‘mobile phone’, *pokrőco* ‘blanket’, *sókro* ‘father-in-law’, *telefóno* ‘telephone’.

The word *telefóno* has apparently got an alternative nominative form, *telefóni*, and there are some other masculine nouns ending in /i/ which show variation, like *kirčimãri* ‘bartender’, *kirãji* ‘king’ and *vitézi* ‘brave warrior’. The fact that we may find variation in the oblique form of lexical items the nominative singular ending of which is *-i* needs further investigation and confirmation from present-day data. The fact that the oblique form of the word *telefóni/telefóno* ‘telephone’, for example, appears both as *telefonés-* and as *telefonós-* might as well be the result of the different nominative forms. Similar instances have been attested, for example the coexistence of *tudōšó* and *tudōší* ‘scientist’ or *mu̍šoró* and *mu̍šorí* ‘programme’. With regard to the variation in *kirčimãri* ‘bartender’, *kirãji* ‘king’ and *vitézi* ‘brave warrior’, we must note that there are ambiguous cases, but

there is not enough information available to draw a conclusion from them. In Cech & Heinschink (1999), we can also find examples from Austrian Lovari where words with a stem-final /i/ have oblique forms only with *-os-*, such as *juhãsi* ‘shepherd’ ~ *juhāsós-*, *doktóri* ‘doctor’ *doktorós-* etc. The Hungarian data, however, do not confirm this, and the regulative nature of the work also casts doubts on this statement.

6.3.2 Possible causes and explanations

6.3.2.1 Variation in the position of stress

In this section, we will look at the relationship between the variation in the position of stress and the appearance of one or the other oblique form and we will see that even though one is not the direct consequence of the other (as the choice of words, *bakró* ‘sheep’ and *sókro* ‘father-in-law’, in Table 31, where the two different patterns are presented, intentionally suggests), there is certainly correlation between the two aspects, which means that there are certain other factors that we might want to take into consideration besides the stem-final vowel.

A possible cause of the variation in the oblique forms, which needs further investigation, is the variation in stress. 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 suffix will be *-es-/-en-*, and where the stress falls on the first (penultimate) syllable, the oblique suffix will be *-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 *bakró* ‘sheep’ will give *bakrés-*. On the level of the word, so on the surface, this results in penultimate stress: dative *bakréske*, locative *bakréste*, ablative *bakréstar* and instrumental *bakrésa*. A child who is acquiring Lovari as their mother tongue can base their assumptions concerning the oblique form on stress in case of disyllabic words.

For words with three syllables, on the other hand, stress may vary widely. While the oblique forms will always have penultimate stress, **the stress of the nominative forms can fall anywhere between the first through the penultimate to the last syllable**. The trisyllabic masculine nouns, along with their oblique forms, are repeated here for

convenience. Table 38 is divided into three sections. The first one contains the nouns with the oblique forms *-es-/-en-*, the second one contains the nouns with the oblique forms *-os-/-on-*, while the third one contains the nouns where there is variation. Their order follows their ending, first the ones with /i/, then the ones with /o/.

word	oblique form
nouns with the oblique form *-es-/-en-*	
bēréši ‘farmhand’	*bērešés-*
bójtāri ‘shepherd boy’	*bojtārés-*
čapóši ‘bartender’	*čapošés-*
farkáši ‘wolf’	*farkašés-*
fillḗri ‘penny’	*fillērés-*
gḗzeši ‘train’	*gēzešés-*
xanrāló ‘policeman’	*xanrālés-*
juhã́si ‘shepherd’	*juhāsés-*
kirã́ji ‘king’	*kirājés-*
kóčiši/kočíši ‘coachman’	*kočišés-*
koldúši ‘beggar’	*koldušés-*
kopã́či ‘tree trunk’	*kopāčés-*
murmúnci ‘cemetery’	*murmuncés-*
pohã́ri ‘glass’	*pohārés-*
téngeri ‘sea’	*tengerés-*
vitḗzi ‘brave warrior’	*vitēzés-*
raklōró ‘little boy’	*raklōrés-*
šārkānʲí ‘dragon’	*šārkānʲés-*
šavōró ‘little boy’	*šavōrés-*
padlṓvo ‘floor’	*padlōvés-*
žamutró ‘son-in-law’	*žamutrés*
nouns with the oblique form *-os-/-on-*	
ã́lato ‘animal’	*álatós-*
bárāto ‘friend’	*barātós-*
bārṓvo ‘baron’	*bārōvós-*
čalã́do ‘family’	*čalādós-*
falató ‘a little bit of food’	*falatós-*

xāmásko ‘food’	*xāmaskós-*
hércego ‘prince’	*hercegós-*
jōsa̋go ‘livestock’	*jōsāgós-*
laptópo ‘laptop’	*laptopós-*
rablő́vo ‘robber, highwayman’	*rablōvós-*
sómsēdo ‘neighbour’	*somsēdós-*
únoko ‘grandchild’	*unokós-*
vonáto ‘train’	*vonatós-*
nouns with variation	
čokáno ‘hammer’	*čokanés-/čokanós-*
dúhano ‘tobacco’	*duhanés-/duhanós-*
mobílo ‘mobile phone’	*mobilés-/mobilós-*
pokrő́co ‘blanket’	*pokrōcés-/pokrōcós-*

Table 38

Trisyllabic masculine nouns and their oblique forms

As we can see, the position of the stress cannot unambiguously predict the oblique form. While it is true that words with stem-final stress take the oblique forms *-es-/-en-* without exception[66], the oblique form of words where the stress shifts to a penultimate or ante-penultimate position is not so obvious. The words *padlő́vo* ‘floor’ and *rablő́vo* ‘robber, highwayman’ have the oblique forms *padlōvés-* and *rablōvós-*, respectively, in spite of the fact that both have penultimate stress, and variation occurs on the level of stems as well, like in the case of *mobílo* ‘mobile phone’ ~ *mobilés-/mobilós-*) or *dúhano* ‘tobacco’ ~ *duhanés-/duhanós-*. The choice of pattern may further be complicated by the fact that the stress of the nominative form may even vary within one stem, as can be seen in the attested example of *kóčiši/kočíši* ‘coachman’. In sum, where stress begins to vary (in words with three or more syllables), the oblique suffix will begin to vary, too.

66 It needs further confirmation, but based on our personal observation it seems that even for the words listed in the tables with stem-final stress, the position of the stress may shift towards the beginning of the word.

6.3.2.2 The number of syllables

There might be a correlation between the number of the syllables a noun has and the degree of variation it shows concerning the oblique forms. This is what we will examine in this section, eventually coming to the conclusion that the higher the number of syllables are, the more likely it is that the oblique form will vary.

Monosyllabic nouns always end in a consonant and invariably take the same oblique pattern, so *drom* 'road' and *dromés-* 'road' obl. This pattern is valid for other nouns ending in a consonant, that are disyllabic, so *rašáj* 'priest' and *rašajés-* 'priest' obl. Variation begins when two factors appear simultaneously: disyllabicity and a stem-final vowel. **A stem-final vowel introduces a certain amount of disturbance in the system, because it conflicts with the initial vowel of the oblique suffix**, which is straightforward for consonant-final stems.[67] However, at that stage, the degree of variation only exists within the category, not on the level of individual paradigms (by category here, we mean the disyllabic stems with a stem-final /o/ – nouns with two syllables and a stem-final /i/ invariably have *-es-/-en-* in their oblique forms). **Every lexical item which has two syllables and a stem-final /o/ will choose either one or the other pattern**, and the position of the stress (final or penultimate) appears to be a reliable clue in this case.

There are 15 disyllabic masculine nouns with a stem-final /o/ and with the oblique forms *-es-/-en-*.

67 The possible special, default or epenthetic nature of /e/ in Lovari has been touched upon.

word	oblique form
ānró ‘egg’	*ānrés-*
bakró ‘sheep’	*bakrés-*
bāló ‘pig’	*bālés-*
bašnó ‘cock’	*bašnés-*
dumó ‘back’	*dumés-*
gāžó ‘non-Romani man’	*gāžés-*
gōnó ‘sack’	*gōnés-*
xānró ‘sword’	*xānrés-*
kurkó ‘week’	*kurkés-*
mānró ‘bread’	*mānrés-*
māšó ‘fish’	*māšés-*
punró ‘foot, leg’	*punrés-*
rakló ‘boy’	*raklés-*
šāvó ‘boy’	*šāvés-*
šēró ‘head’	*šērés-*

Table 39

Disyllabic masculine nouns with a stem-final /o/ and with the oblique form *-es-/-en-*

There are 19 disyllabic masculine stems with a stem-final /o/ and *-os-/-on-* as their oblique forms.

When the number of syllables rises to three, variation begins on the level of lexical items (it seems to be both intra- and inter-speaker variation). This would mean that the longer a word is, the more uncertain it gets which oblique stem it will take. There is only slight variation for words longer than two syllables which end in a different vowel, like /i/: the frontness of the stem-final vowel will dominantly predict (or trigger) a front vowel in the oblique form. The back vowel /o/ of nouns with three syllables, however, will not be able to predict the oblique form unambiguously, just like disyllabic nouns ending in /o/ cannot.

noun	oblique form
ā*tko* ‘curse’	*ātkós-*
bógo ‘nag’	*bogós-*
búso ‘bus’	*busós-*
čā́so ‘hour, watch’	*čāsós-*
djā́so ‘mourning’	*djāsós-*
fő́ro ‘town’	*fōrós-*
gíndo ‘problem’	*gindós-*
grő́fo ‘count’	*grōfós-*
hī́ro ‘a piece of news’	*hīrós-*
nā́so ‘child’s father-in-law’	*nāsós-*
nī́po ‘relatives’	*nīpós-*
pápo ‘grandfather’	*papós-*
pā́to ‘bed’	*pātós-*
pérco ‘minute’	*percós-*
práho ‘dust’	*prahós-*
púju ‘chicken’	*pujós-*
ríto ‘field’	*ritós-*
sóbro ‘statue’	*sobrós-*
sókro ‘father-in-law’	*sokrós-*

Table 40

Disyllabic masculine stems with a stem-final /o/ and *-os-/-on-* as their oblique form

We checked the words with three syllables with a stem-final /o/ in the newly collected data and in our other source of Hungarian Lovari, Vekerdi (1985) for variation. We found the lexical items in Table 41 which had oblique occurrences as well (either singular or plural or both).

word	oblique form	variation
nouns with the oblique form *-es-/-en-*		
xanrāló ‘policeman’	*xanrālés-*	no
padlṓvo ‘floor’	*padlōvés-*	no
raklōró ‘little boy’	*raklōrés-*	no
šavōró ‘little boy’	*šavōrés-*	no
žamutró ‘son-in-law’	*žamutrés-*	no
nouns with the oblique form *-es-/-en-*		
ãlato ‘animal’	*álatós-*	no
bárāto ‘friend’	*barātós-*	no
bārṓvo ‘baron’	*bārōvós-*	no
čálādo ‘family’	*čalādós-*	no
falató ‘a little bit of food’	*falatós-*	no
xāmásko ‘food’	*xāmaskós-*	no
hércego ‘prince’	*hercegós-*	no
jōsãgo ‘livestock’	*jōsāgós-*	no
laptópo ‘laptop’	*laptopós-*	no
rablṓvo ‘robber, highwayman’	*rablōvós-*	no
sómsēdo ‘neighbour’	*somsēdós-*	no
únoko ‘grandchild’	*unokós-*	no
vonáto ‘train’	*vonatós-*	no
nouns with both oblique forms		
čokáno ‘hammer’	*čokanés-/čokanós-*	yes
dúhano ‘tobacco’	*duhanés-/duhanós-*	yes
mobílo ‘mobile phone’	*mobilés-/mobilós-*	yes
pokrṓco ‘blanket’	*pokrōcés-/pokrōcós-*	yes

Table 41

Trisyllabic masculine stems with a stem-final /o/ and their oblique forms

Among the trisyllabic masculine nouns[68] with a stem-final /o/, there are only 5 which take *-es-/-en-* as their oblique as opposed to 13 with the oblique form *-os-/-on-*. In addition, there are 4 stems whose oblique forms vary. This is somewhat in line with the varying stress pattern of trisyllabic nouns: **the increase in the number of syllables increases the chance of variation**, too. While the oblique form of disyllabic nouns never varies (it is either *-es-/-en-* or *-os-/-on-*), it is fairly conspicuous that when the number of syllables exceeds two, the oblique form begins to vary. It should also be noted here in connection with the higher number of *-os-/-on-* oblique forms 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 form which contains an /o/.

6.3.2.3 The plural form

In this section, we will demonstrate how **the nominative plural form of nouns relates to both the singular and plural oblique forms**. Based on the data at hand, we will see that there is **a close relationship** between the two, but one of the two possible plural endings will not predict the oblique form precisely.

There are two possible candidates for the ending of the plural form of masculine nouns with a stem-final /o/ (see also section 5.1.4.2). Both suffixes truncate the stem, so the /o/ is deleted. One of them is the suffix *-e*, the other one is the suffix *-ura*. If the suffix is /e/, there is no variation in the oblique forms at all, at least no variation has been attested so far, as we can see from the examples in Table 42.

68 Although there are some four-syllable words, too, the reason why we did not take them into consideration is their ambiguous status: they may only be ad hoc borrowings. The extent to which borrowing from Hungarian takes place “online” in the case of individual lexical items is of course not possible to tell exactly, but we decided not to take into account words which only appear within the speech of one speaker and even then only once (or twice, including the nominative form).

plural form	attested oblique forms
bālé ‘pig’ pl.	*bālés*, *bāléske*, *bālésa*, *bālénca* etc.
raklé ‘boy’ pl.	*raklés*, *rakléski*, *raklésko*, *rakléske* etc.
šēré ‘head’ pl.	*šēréngo*, *šēréste* etc.
xanrālé ‘policeman’ pl.	*xanrālén*

Table 42

Examples of masculine nouns with the plural suffix *-e* and their oblique forms

This comes as no surprise: the plural form ending in /e/, more precisely the front vowel itself can easily trigger the oblique suffixes *-es-* and *-en-*, which also contain a front vowel. The singular form, ending in /o/ provides no phonological clue as to the nature of the oblique forms. However, the uncertainty is reduced by the plural form.

The plural suffix *-ura*, just like the singular ending *-o*, provides no obvious clue regarding the oblique forms.[69] We see variation, because the relationship is far from being so clear-cut as in the case of the plural suffix *-e*. As for the data, we used Vekerdi (1985) and the newly collected data. Based on the data in Table 43 it seems that the plural suffix *-ura* (*-uri*, *-ure*) itself cannot determine the oblique forms. The situation is further complicated by the fact that it is not possible to predict the nominative singular from the plural: the nominative singular ending can be both *-o* and *-i*; in addition, we find a few cases where the nominative singular turns out to be a feminine noun, like in the case of *grifmadǎra* ‘gryphon’. The contents of Table 43 are arranged according to the number of syllables of the nominative singular again.

69 The suffix *-ura* appears in Vekerdi (1985) in the forms *-uri* and *-ure*, too. The form *-uri* is the form borrowed from Romanian, but both *-ure* and *-ura* seem to be inner innovations triggered by the other two typical nominative masculine singular forms ending in /e/, like *bakró/bakré* ‘sheep’, and /a/, like *rom/romá* ‘man’.

plural form	attested oblique forms	variation	nominative singular
two syllables			
bájura ‘problem’ pl.	–	?	*bájo*
čãsura ‘hour’ pl.	*čāsónde*	no	*čãso*
fő́rura ‘town’ pl.	*fōrós*	no	*fő́ro*
gínduri/gíndura ‘problem’ pl.	*gindóstar, gindónca*	no	*gíndo*
gī́pura ‘computer’ pl.	*gīpósa*	no	*gī́po*
grő́fura ‘count’ pl.	*grōfóske*	no	*grő́fo*
hī́rura ‘a piece of news’ pl.	*hīróstar*	no	*hī́ro*
nī́puri/nī́pura ‘relative’ pl.	*nīpós*	no	*nī́po*
sóbruri ‘statue’ pl.	*sobrós, sobrónde, sobróske*	no	*sóbro*
trájura ‘life’ pl.	–	?	*trájo*
three syllables			
abrónčura ‘tyre’ pl.	–	?	*abrónči*
ālatúri ‘animal’ pl.	*ālatónge, ālatón, ālatós*	no	*ãláto*
barātúra ‘friend’ pl.	*barātóske*	no	*bárāto*
bojtārúra ‘shepherd boy’ pl.	*bojtārénge*	no	*bojtãri*
čokanúra ‘hammer’ pl.	*čokanésko, čokanósko*	yes	*čókano*
felhő̄vúri	–	?	*felhő̄vo*
juhãsura ‘shepherd’ pl.	*juhāséske, juhāsés, juhāséstar*	no	*juhãsi*
láptopura ‘laptop’ pl.	*laptopósa*	no	*láptopo*
móbilura ‘mobile phone’ pl.	*mobilésa, mobilósa*	yes	*móbilo*
vítēzuri ‘brave warrior’ pl.	*vitēzós, vitēzón, vitēzés, vitēzéske*	yes	*vítēzi*
five syllables			
grifmadãrure ‘gryphon’ pl.	*grifmadāráke*	no	*grifmadãra*

Table 43

Nouns with the plural form *-ura* (*-uri*, *-ure*)

We find 2 lexical items which take the plural suffix *-ur(i/a/e)* and the oblique *-es-/-en-*. The number of those which have *-os-/-on-* as their oblique form is 11. One word turns out to inflect according to the feminine paradigm, as mentioned above. There is variation in case of 3 items and 4 nouns did not have attested oblique forms. All in all, we can conclude

that the plural suffix *-e* will predict the oblique form *-es-/-en-*, while the plural suffix *-ur(i/a/e)* will predict the oblique form *-os-/-on-* with fairly high certainty. This is demonstrated in Figure 10, where the schemata for the oblique and the plural are connected through dashed arrows, the thickness of which represents the likelihood with which the plural forms will predict the oblique forms.

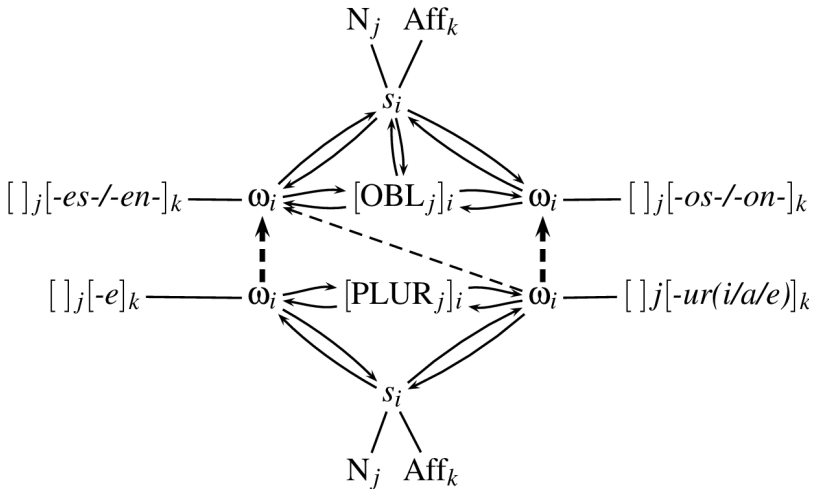

Figure 10

Schemata for the masculine oblique ending and the masculine plural ending

6.3.2.4 The masculine adjectival ending -ano

The reason why the masculine adjectival ending *-ano*[70] might be interesting in connection with the variation we see is **its similarity to the ending of the nouns demonstrating the most variation** *čókano* 'hammer' and *dúhano* 'tobacco'. The feminine ending *-ani* is not relevant, but the plural and the oblique ending *-ane* is relevant, too, due to the vowel alternation in *-ano* ~ *-ane*. In the newly collected data and Vekerdi (1985) we found 18 adjectives altogether; these can be seen in (29).

(29)*bakranó* 'of sheep'
balanó 'of pigs'
čačikanó 'real'
čiríkjanó 'of birds'

70 Also in the forms *-kano* and *-ikano*.

daranó ‘fearful, frightened’
dulmutanó ‘old, former’
džuklanó ‘of dogs’
gažikanó ‘non-Gypsy, non-Romani’
grastanó ‘of horses’
guruvanó ‘of cattle’
kaštanó ‘wooden’
manušanó ‘human’
mulikanó ‘of death’
puranó ‘old, former’
rajkanó ‘gentleman-like’
romanó ‘Gypsy, Romani’
šošojanó ‘of rabbits’
žuljanó ‘of women’

They are mainly used as attributive adjectives, as in *grastanó bal* ‘horse hair’, *guruvanó mas* ‘beef’, *muró dulmutanó vurdón* ‘my old car’, *gažikané gādá* ‘non-Gypsy clothes’ (plural!), *manušanó rat* ‘human blood’ etc. If the nominal head of the phrase is inflected, the adjectives take on their oblique form: instrumental *grastané balésa*, *manušané ratésa*, dative *guruvané maséske*, ablative *muré dulmutané vurdonéstar* etc. **These forms can influence the oblique form of nouns**, as they are *-ano* in the nominative singular and *-ane* in the oblique, so for example the instrumental of the phrase *kaštanó čókano* ‘wooden hammer’ could be both *kaštané čókanésa* and *kaštané čókanósa*, and we suppose that the former one is more likely.

In addition, the adverbs derived from these adjectives take the ending *-anes*, which corresponds exactly to the words *čokanés* and *duhanés*: *romanés* ‘in a Romani way’, *čačikanés* ‘really, in actual fact’.

The situation is not so straightforward, however, as, according to Vekerdi (1985, 2000), the length of the /a/ in the masculine adjectival endings appears to vary in Vlax Romani varieties, like Lovari, so *dulmutānó* instead of *dulmutanó*, *čiríkjānó* instead of *čirikjanó*, *šošojānó* instead of *šošojanó*, *žuljānó* instead of *žuljanó* etc. However, this may not only be true for the adjectives. The excerpt in (30) from the Romungro variety shows

that the length of the vowel varies within the speech of the same speaker (Vekerdi 1985: 296).

(30) *Šaj* *sīvináv,* *phenél,* *duj*
‘can’ MOD. AUX. POS. ‘smoke’ 1ST SING. PRES. IND. ‘say’ 3RD SING. PRES. IND. ‘two’
gōnó *dúhāno?* *Naštíg*
‘sack’ ACC. SING. ‘tobacco’ ACC. SING. ‘can’ MOD. AUX. NEG.
sīvinés, *phenél,* *čak* *epáš*
‘smoke’ 2ND SING. PRES. IND. ‘say’ 3RD SING. PRES. IND. ‘only’ ‘half’
duhanéha; *epáš* *gonéha.*
‘TOBACCO’ SING. INSTR. ‘half’ ‘sack’ SING. INSTR.
‘“Can I smoke,” he says, “two sacks of tobacco?” “You can’t,” he says, “only two sacks.”’

The long /a/ in *dúhāno* becomes short in *duhanéha*, and the same happens to the long /o/ of *gōnó* in *gonéha*. Similar phenomena can happen in other varieties (for example, Vekerdi (2000) explicitly says that *romānés* ‘in a Gypsy/Romani way’ has got a long vowel, whereas the newly collected data show that this is not the case, and, actually, *romanés* with a short vowel is more common). Therefore, **it is possible to claim that there is correlation between the variation of oblique forms and the masculine adjectival endings** ***-ano/-ane***: these endings are likely to support the appearance of the oblique endings *-es-/-en-* for masculine nouns with a stem final /o/. This is shown in Figure 11 through the relationship between the schema for the oblique and the schema for the *-ano/-ane*, represented by the one thick dashed arrow leading from the phonological form in the latter to the oblique endings *-es-/-en-*, rather than the oblique endings *-os-/-on-*.

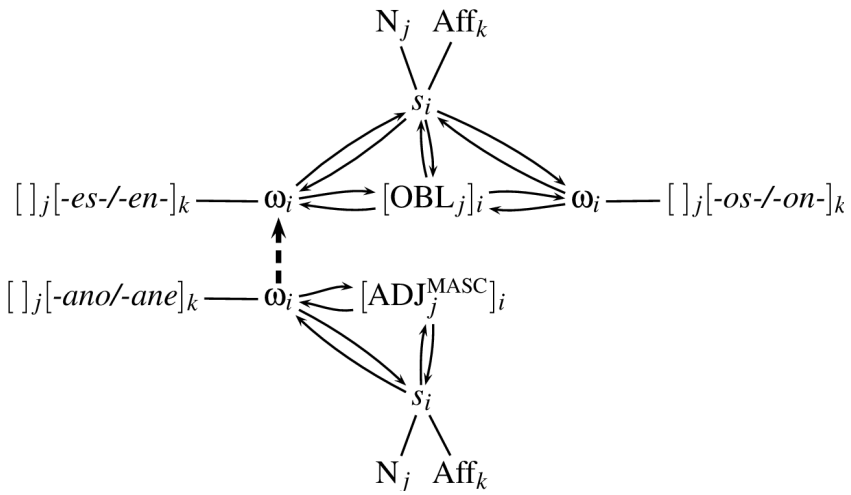

Figure 11

The relationship between the schema for the masculine oblique ending and the adjectival ending *-ano/-ane*

6.3.2.5 2nd person singular verbal endings

Although verbs constitute a different part of speech, on the level of surface forms the possibility of the influence of forms with an identical ending cannot be completely excluded. The endings under discussion are the nominal endings -es and -*os*, and both endings appears on verbs, too. As we could see in section 5.2.2, the 2nd person singular present indicative of the most numerous class, the consonantal verbs, ends in -*es*: *beš-* 'sit, live' ~ *bešés*. Its effect is best measured against the 2nd person singular present indicative ending -*os* appearing on mediopassive verbs: *paš^{j}uv-* 'lie' ~ *paš^{j}ós*.[71] Their frequency is not to be underestimated, either, especially if we consider their overall frequency. It is also worth noting that these personal concord markers are invariably stressed, just like the oblique endings of nouns. The verbs with -*es* as their 2nd person singular present indicative marker (in other words, the members of the consonantal class) attested in the newly collected data and Vekerdi (1985) can be seen in Table 44. Although we rely on type frequency here, in order to provide some extra information, we also added in which set of data they were attested and whether the 2nd person singular present indicative form was to

71 In addition, there are some other, primarily borrowed verbs whose 2nd person singular present indicative forms take the personal concord marker -*os*. Although their total number is low, if they are attested in the newly collected data, they are also included in the table.

be found in particular. Table 44 contains 103 lexical items in total.

verb	new data	Vekerdi (1985)	2nd person singular present indicative
akʰar- ‘call, summon’	no	yes	no
akʰarav- ‘invite, send for’	no	yes	no
an- ‘bring’	yes	yes	no
astar- ‘take, catch’	yes	yes	no
aš- ‘remain’	no	yes	yes
av- ‘come’	yes	yes	yes
barvar- ‘make rich’	no	yes	no
bašav- ‘play music’	yes	no	yes
beš- ‘sit, live’	yes	yes	yes
bišal-/bišav- ‘send’	yes	no	no
biťʲin- ‘sell’	yes	yes	yes
bold- ‘turn’	no	yes	no
brist- ‘forget’	yes	yes	yes
cipin- ‘shout’	yes	no	no
cird- ‘pull’	yes	yes	no
colaxar- ‘get married’	no	yes	no
čor-/čōr- ‘steal’	yes	yes	no
d- ‘give’	yes	yes	yes
darav- ‘frighten’	no	yes	yes
dikʰ- ‘see, look’	yes	yes	yes
dijajv- ‘go mad’	yes	no	no
diljār- ‘drive mad’	no	yes	yes
ďʲinav- ‘read’	no	yes	no
drabar- ‘tell fortune’	no	yes	no
emlēksin- ‘remember’	yes	no	yes
ērkezin- ‘arrive’	yes	no	no
garuv- ‘hide’	no	yes	no
xaxav- ‘feed’	yes	yes	no
xasajv- ‘disappear, get lost’	no	yes	no
haťʲār- ‘understand’	yes	yes	no

hohav-/xoxav- ‘lie’	yes	yes	yes
xojajv- ‘get angry’	yes	yes	no
xojar- ‘make angry’	yes	no	yes
xunav- ‘dig’	yes	no	no
xut^{j}il- ‘take, seize’	yes	yes	no
xutt^{j}- ‘jump’	no	yes	no
ingr- ‘carry’	yes	yes	yes
inkr- ‘hold’	yes	yes	yes
kam- ‘love’	yes	yes	yes
ker- ‘make, do’	yes	yes	yes
kerav- ‘have something made/done’	yes	yes	yes
k^{h}el- ‘play, dance’	yes	yes	yes
khos- ‘wipe’	yes	no	no
kin- ‘buy’	no	yes	no
kirav- ‘cook’	no	yes	no
korrajv- ‘go blind’	yes	no	no
kur(r)- ‘fuck’	no	yes	no
l- ‘take’	yes	yes	yes
lamint- ‘watch, keep an eye on’	yes	no	no
lašar- ‘make better’	no	yes	no
malaptatisar- ‘have (a horse) shod’	no	yes	no
malav- ‘hit, strike’	yes	no	no
mang- ‘ask’	yes	yes	yes
mangav- ‘propose to’	yes	no	no
mār- ‘beat’	yes	yes	yes
mēr- ‘die’	no	yes	yes
mudār- ‘kill’	no	yes	yes
muk(^{h}*)-* ‘let, leave’	yes	yes	yes
murdajv- ‘die’	yes	yes	no
naš- ‘run’	yes	yes	no
našav- ‘chase away’	no	yes	no
paruv- ‘change’	yes	no	yes
pat^{j}ār- ‘pack up’	yes	no	no

per-/pēr- ‘fall’	yes	yes	no
phabār- ‘set fire to’	no	yes	no
phag- ‘break’	yes	yes	no
phand- ‘bind, tie’	yes	yes	no
phen- ‘say’	yes	yes	yes
p^{h}ir/p^{h}īr- ‘walk, go’	yes	yes	yes
phirav- ‘wear’	yes	no	no
phurd- ‘blow’	yes	yes	no
pinžār- ‘know’	yes	yes	no
pizd- ‘push’	no	yes	no
pot^{j}in- ‘pay’	no	yes	no
puš- ‘ask’	no	yes	no
pusav- ‘stab, prick’	yes	yes	no
putr- ‘open’	no	yes	no
rak^{h}- ‘find’	yes	yes	no
randr- ‘undress’	no	yes	no
res- ‘reach’	yes	yes	yes
rod- ‘look for’	yes	yes	yes
rov- ‘cry’	yes	yes	no
sast^{j}ār- ‘heal, cure’	yes	yes	no
sid^{j}ār- ‘hurry’	yes	yes	no
sikav- ‘show’	no	yes	no
sit^{j}ār- ‘teach’	yes	no	no
šarav- ‘cover’	yes	no	no
šin- ‘cut’	yes	yes	yes
šingr- ‘tear’	yes	yes	yes
šor- ‘pour’	yes	yes	yes
šud- ‘dob’	no	yes	no
sōv- ‘sleep’	yes	yes	yes
tagad- ‘deny’	yes	no	yes
thov- ‘wash’	yes	yes	no
tord^{j}ār- ‘stop’	no	yes	no
trad- ‘drive’	no	yes	no
urav- ‘dress’	yes	no	no

ušťʲav- ‘wake up, step’	yes	yes	no
vazd- ‘lift’	no	yes	no
zumav- ‘try’	no	yes	no
zurajv- ‘become strong’	yes	no	yes
žan- ‘know, can’	yes	yes	yes
žuťʲār- ‘wait’	yes	no	no

Table 44

Consonantal verbs (2nd person singular present indicative marker *-es*)

In Table 45, we can see the verbs with *-os* as their 2nd person singular present indicative marker, in the same manner as above, containing only items which were attested in either the newly collected data or in Vekerdi (1985).

verb	new data	Vekerdi (1985)	2nd person singular present indicative
arakhadʲuv- ‘be born’	no	yes	no
ašadʲuv- ‘fall’	yes	no	no
bāruv- ‘grow’	yes	no	yes
bašuv- ‘give sound’	no	yes	no
bušuv- ‘be called’	no	yes	yes
čhinʲuv- ‘become tired’	yes	no	no
dičhuv-/dičjuv- ‘be seen’	yes	yes	no
džuvindjuv- ‘become alive’	no	yes	no
ērto- ‘forgive’	yes	yes	yes
garadʲuv- ‘hide’	yes	no	yes
gindo- ‘think’	yes	no	yes
kalʲuv- ‘become black’	no	yes	no
kerdjuv- ‘be born, become’	yes	yes	no
maladʲuv- ‘meet’	yes	yes	yes
načuv- ‘pass’ intr.	no	yes	no
najuv- ‘have a bath’	yes	no	no
māťʲuv- ‘get drunk’	no	yes	no

paho- ‘catch a cold’	yes	no	no
paš^{j}uv- ‘lie’	yes	no	yes
phabuv- ‘burn’	no	yes	no
phadjuv-/phagerdjuv- ‘break’ intr.	yes	no	no
phūruv- ‘get old’	yes	no	no
puterd^{j}uv- ‘open’	no	yes	no
resad^{j}uv- ‘meet’	yes	no	yes
sero- ‘remember’	yes	no	yes
sit^{j}uv- ‘learn’	yes	no	yes
šindjuv- ‘get torn’	yes	no	no
šordjuv- ‘flow’	yes	no	no
šuv- ‘put’	yes	yes	yes
ternjuv- ‘become young’	yes	no	no
t^{j}ino- ‘shake’	yes	no	no
t^{j}iruv- ‘cook’ intr.	yes	no	no
t$^{j(h)}$išjuv- ‘lose weight’	yes	no	yes
tord^{j}uv- ‘stand’	yes	yes	yes

Table 45

Mediopassive verbs (2nd person singular present indicative marker *-os*)

This table contains 33 items, so the proportion of the consonantal verbs (those with *-es* as their 2nd person singular present indicative marker) and the mediopassive verbs (those with *-os* as their 2nd person singular present indicative marker) is 3.1:1. If we look back at the masculine nouns with the oblique marker *-es-/-en-* and the masculine nouns with the oblique marker *-os-/-on-* and calculate their proportion, it is 2.3:1 (not including the stems which vary). If we compare the two proportions, we can see that in both cases, the forms with the vowel /e/ are in majority. On the one hand, this might reflect the influence of the verbal system, more specifically the 2nd person singular forms of the verbs on the oblique forms of masculine nouns. On the other, as the proportions show a similar tendency but do not exactly correspond (~3:1 and ~2:1), it might reflect a more general tendency in the language, preferring front vowels to back vowels in suffixes.

The influence of the 2nd person singular forms on the singular oblique forms of

masculine nouns is shown in Figure 12, where the two schemata are connected again with thick dashed arrows. **The 2nd person singular form *-es* will tip the scales in favour of the oblique ending *-es-*, while the 2nd person singular form *-os* will do the same in favour of the oblique ending *-os-*.**

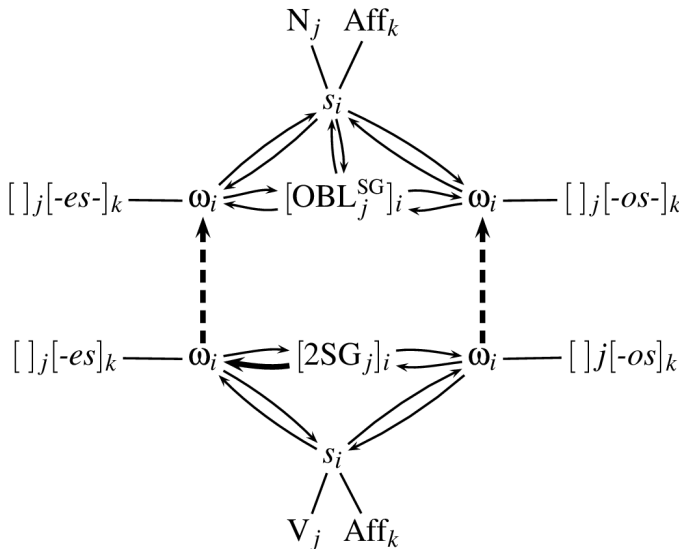

Figure 12

The relationship between the schema for the masculine singular oblique ending and the 2nd person singular form of consonantal and mediopassive verbs

6.3.2.6 The adverbial ending -es

As mentioned in section 6.3.2.4, adverbs derived from adjectives ending in *-ano/-ani/-ane* end in *-anes*, which corresponds exactly to the masculine singular oblique ending of words like *čókano* and *dúhano*: *čokanés* and *duhanés*. Other adverbs derived from adjectives also end in *-es*, which corresponds to the masculine singular oblique ending *-es.* **Their use and frequency might play a role in a bias towards *-es-* as the oblique marker instead of *-os-*.**

We can get some idea about their usage and their syntactic position from examples (31)-(36). The first one is a very common question, the second one stems from an Internet comment, the third to the sixth come from the newly collected data.

(31)*Žanés* ***romanés****?*

‘know’ 2ND SING. PRES. IND. ‘in the Romani way’ ADV.

‘Do you speak Romani?’

(32)***Čačikanés*** *kamáv* *tut.*

‘really, in actual fact’ adv. ‘love’ 1ST SING. PRES. IND. 2ND SING. PERS. PRON. ACC.

‘I really love you.’

(33)*Bužangló* *kraj* *sim,*

‘cunning’ MASC. SING. ‘king’ SING. NOM. ‘be’ 1ST SING. PRES. IND.

bužanglés *trajíj.*

‘cunningly’ ADV. ‘live’ 1st SING. PRES. IND.

‘I am a cunning king, I live cunningly.’

(34)*Godjavér* *kirãji* *sim,*

‘clever’ MASC. SING. ‘king’ SING. NOM. ‘be’ 1ST SING. PRES. IND.

godjaverés *uralkodíj.*

‘cleverly’ ADV. ‘reign’ 1st SING. PRES. IND.

‘I am a clever king, I reign cleverly.’

(35)***Rossulés*** *kérdjilem* *kathár*

‘badly’ ADV. ‘become’ 1ST SING. PAST IND. ‘from’ PREP.

o *hĩro.*

ART. DEF. SING. MASC. NOM. ‘a piece of news’ SING. NOM.

‘The news made me feel ill.’

(36)***Nāsulés*** *simás* *e*

‘badly’ ADV. ‘be’ 1ST SING. PAST IND. ART. DEF. SING. MASC. OBL.

hīróstar.

‘a piece of news’ SING. ABL.

‘The news made me feel ill.’

In the newly collected data and in Vekerdi (1985) we found 12 such adverbs in total. They are listed in Table 46, along with the adjective they are derived from.

adverb	adjective
baxtālés ‘happily, luckily’	*baxtāló*
boldinés ‘upside down’	*boldinó*
bužanglés ‘cunningly’	*bužangló*
čačikanés ‘really’	*čačikanó*
čordānés ‘secretly’	*čordānó*
godjaverés ‘cleverly’	*godjavér*
godjaverikānés ‘cleverly’	*godjaverikānó*
mātés ‘drunkenly’	*mātó*
nāsulés ‘badly’	*nāsúl*
romanés ‘in Romani, in a Romani way’	*romanó*
rossulés[72] ‘badly’	–
šukārés ‘nicely, gently’	*šukā̆r*

Table 46

Adverbs ending in *-es* and the adjectives they are derived from

At present, the possible analogical effect is merely based on the identical phonological form of the adverbial ending and the oblique ending *-es-*. This effect is shown in the relationship of the oblique schema and the adverbial schema in Figure 13. The thick dashed arrow means that the adverbial ending *-es* would tip the scales in favour of the oblique ending *-es-*.

72 This is a fine example of a different kind of analogical phenomenon: the schema containing the derivational affix *-es* in the phonological and the adverbial function in the semantic component begins to be applied in an apparently unusual way, partly ignoring the syntactic component which should include an adjective.

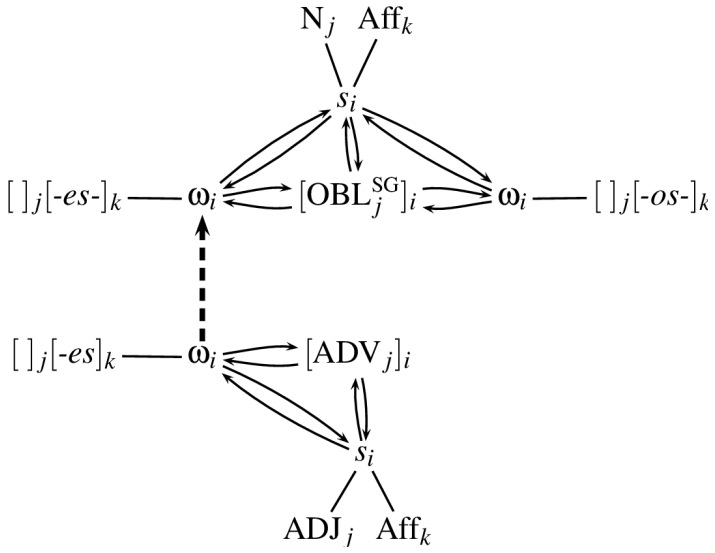

Figure 13

The relationship between the masculine singular oblique endings and the adverbial ending *-es* as represented by schemata

In order to have more convincing evidence of the possible effect, we will need to calculate token frequency, too, in a larger corpus.

6.3.3 Conclusion

In this section, we had a look at the first weak point in Lovari morphology, the masculine oblique base, in more detail. Following the description of the phenomenon in question, we went over six possible reasons for the weakness and the ensuing variation 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, too, so **the varying stress pattern pairs up with the unpredictability of oblique forms**.

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.
3. The plural form. There are two possible nominative plural endings for masculine nouns. It seems that **the plural ending can provide us with some clue as to the distribution of the oblique ending**, but it can be predicted only partially. In addition, we must also note that the nominative plural form cannot be predicted unambiguously from the nominative singular.
4. The masculine adjectival ending *-ano*. We saw that there is a set of denominal adjectives whose ending is *-ano* in the nominative and *-ane* in the oblique. These adjectives were found interesting because their oblique form ends in /e/, while their nominative form ends in /o/, which is similar to what we find in one of the patterns for the masculine oblique base, where nouns ending in /o/ take the oblique forms *-es-/-en-*. Although calculations of token frequency are not available at present, based on the evidence we have, we can say that **the *-ano* ~ *-ane* endings support the appearance of the oblique endings *-es-/-en-* for masculine nouns with a stem final /o/**.
5. 2nd person singular verbal endings. We examined the 2nd person singular present indicative ending of verbs (*-es* for those with a stem-final consonant and *-os* for the mediopassive verbs) and the possbile correlation between the proportion of the type frequency of the consonantal verbs and the mediopassive verbs. After comparing their proportions to that of the oblique forms *-es-/-en-* and *-os-/-on-*, we found that **the forms with the vowel /e/ are in majority, which reflects the influence of the verbal system**, more specifically the 2nd person singular forms of the verbs on the oblique forms of masculine nouns.
6. The adverbial ending *-es*. We briefly touched upon the **adverbs mostly derived from adjectives and ending in *-es* and their possible influence on the choice of the singular oblique ending in the masculine in favour of *-es-***, as opposed to *-os-*. The data here are particularly scarce, so we came to the conclusion that this particular aspect needs more data and calculations.

6.3.4 A brief sidetrack: the “inherited-borrowed dichotomy”

We must mention here that in connection with the two different patterns, many (e.g. Boretzky 1989, Bakker 1997, Matras 2002) emphasise the existence of a strict morphological split between the vocabulary inherited from Indo-Aryan (as well as words borrowed from Persian and Armenian) and the vocabulary borrowed later from Greek and other (Romanian, Serbian, Hungarian etc.) contact 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.[73] (Elšík 2000: 17)

In the nominal inflection this would appear like this: one of the patterns (the oblique in *-es-/-en-*) is used to inflect inherited nouns due to historical reasons, the other pattern (the oblique in *-os-/-on-*), being itself borrowed from Greek (Bakker 1997), is used to inflect borrowed nouns. Descriptions of Lovari (Hutterer & Mészáros 1967, Cech & Heinschink 1999) go along this path, with minor differences, so even masculine nouns with a stem-final *-i* take the oblique in *-os-/-on-* (Cech & Heinschink 1999: 22), which is clearly not the case, as we saw in section 6.3. Elšík (2000) discusses the historical development of nominal paradigms in detail, and, regarding the Greek-derived word *fōro* ‘town’, he states that diachronically *fōrós-* replaced *fōrés-*, so that the oblique form could resemble the nominative singular. However, even in a diachronic sense, this is hard to justify, as it goes against the basic layout of the inherited inflection, where the oblique singular stem ends in *-es-*, no matter what the nominative ending is (for example nominative singular *bāló* ‘pig’ and obl. sing. *bālés-*).

We cannot regard the existing situation a result of historical processes, as what we see is variation, not one state or another. Psycholinguistic factors might interfere in the form of the extent to which a native speaker “feels” that a certain word is borrowed or not, but this is very difficult to measure. Intuitively, one would think that, although the word *dúhano* is

73 The terms “thematic” and “athematic” are very misleadingly used instead of “inherited” and “borrowed” in papers focussing on Romani linguistics.

an earlier loan from Serbian than the word *čókano* from Romanian, the similarity of Hungarian *dohány* might evoke a sense of the word being less old.[74] The fact that there is only slight hesitation concerning the oblique forms of *sókro* does not really justify this as the current speakers of Hungarian Lovari have no access to Romanian at all.

All in all, we have to dismiss the notion of the strict inherited-borrowed dichotomy, and thus, its erosion and any 'interaction' (Elšík 2000: 23) between the two layers, too. The two layers do not exist as there are no two specific and unique morphological systems used for one and the other; their inflection, strictly taken, is not different. What we must see clearly is that there are two patterns existing within the masculine paradigm of nouns ending in *-o*, and the choice may depend on several factors, including the overall frequency of the patterns. It is also true that the predominance of *-os-/-on-* forms in the case of *sókro*, for example, can be the result of the frequency of the forms of the particular paradigm itself (token frequency applied to paradigms), like in the case of the paradigm of *fő́ro* 'town', where high token frequency may be the reason for the apparent lack of variation. On the other hand, variation in the case of the oblique form of a word like *čókano* 'hammer' can be the result of its lower token frequency. Other cognitive processes might play a role, too. For example, the extent of embeddedness is difficult to measure, but it may consist of such factors as how deeply embedded the word is mentally in language use, or what other notions might come into play, like even intuitions concerning the "Gypsyness" of the word.

6.4 The feminine oblique plural base

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

1. The masculine oblique plural *-en-*. Besides *-an-*, the other variant of the feminine

74 The plural suffix *-ura*, for example, never seems to attach to inherited words and "earlier" loans, so the plural form of such words as *rom* 'man' and *drom* 'road' (from Greek) are *roma* and *droma*, respectively (Matras 2002: 81). We must notice, however, the phonological similarity of these two words.

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 the influence exists.
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 ending of the nouns with the plural ending *-i* is *-a* and that of the nouns with the plural ending *-a* is often *-i*.

6.4.1 Description of the phenomenon

We have seen in sections 5.1.1 and 5.1.3 that the feminine oblique plural base has got one form in *-an-*, so the oblique base of a word like *krangá* 'branch' is *krangán-* (Hutterer & Mészáros 1967: 49). The Lovari data, however, show variation again: **the forms *-en-* and *-an-* occur simultaneously as the feminine oblique plural base on several points of the feminine paradigm**.[75] This suggests that we are dealing with two competing patterns again.[76]

Table 47 shows 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-*.

75 The word taken as an example, *kranga* 'branch' exclusively inflects as *krangén-* in the oblique, contrary to what is stated by Hutterer & Mészáros (1967).

76 According to the literature (Matras 2002: 83, Elšík 2000: 22, Boretzky 1994: 33), the form *-an-* is the result of a renewal or assimilation on the basis of the oblique singular, *krangá-*, from an original oblique plural in *-en-*, and it most often happens in the Vlax dialects. This suggests that the variation we see here might be a sign of an ongoing change.

feminine	*rā̆ca* ‘duck’		*māčí* ‘fly’	
	singular	plural	singular	plural
N	rā̆ca	rācí	māčí	mā̆čá
A	rācá	rācén	mā̆čá	mā̆čán
D	rācáke	rācénge	mā̆čáke	mā̆čánge
L	rācáte	rācénde	mā̆čáte	mā̆čánde
Abl	rācátar	rācéndar	mā̆čátar	mā̆čándar
I	rācása	rācénca	mā̆čása	mā̆čánca
G	rācák-	rācéng-	mā̆čák-	mā̆čáng-
V	rā̆ca	rācále	mā̆̆ča	mā̆čále

Table 47

The two different patterns in the feminine

The two different patterns can be represented by the following combination of two schemata, shown in Figure 14, 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, $\omega_i[\ ]_j[en]_k$. We can also look at it from the other direction: 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, $\omega_i[\ ]_j[an]_k$.

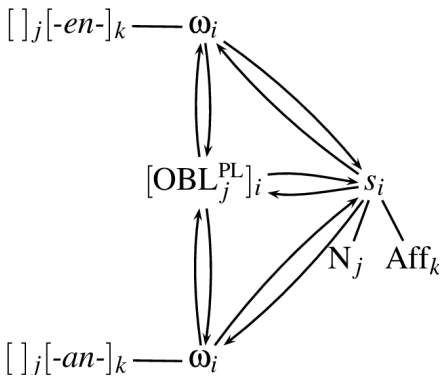

Figure 14

The combination of two schemata for the feminine oblique plural

Table 48 sums up the feminine nouns based on Vekerdi (1985). Only words with at least one attested oblique plural form are included in the list, and they are grouped together according to their oblique plural form. The list contains 16 lexical items. The oblique plural is *-an-* for 7 words, *-en-* for 8 words, and there is one stem which varies.

noun	attested oblique forms
nouns with the oblique plural *-an-*	
papín ‘goose’	*papinʲán/papinʲánge*
phenʲ ‘sister’	*phenʲánca*
raklʲí ‘girl’	*raklʲánge/raklʲánca/raklʲán*
rolʲí ‘stick’	*rolʲánge*
romnʲí ‘woman’	*romnʲánge*
šej ‘girl’	*šejánge*
žuvljí ‘woman’	*žuvlján*
nouns with the oblique plural *-en-*	
birulʲí ‘bee’	*birulʲén*
čerháj ‘star’	*čerhajéngi*
čiriklí ‘bird’	*čiriklʲén*
cóxa ‘skirt’	*coxéngi*
pápuča ‘shoe’	*papučénge*
phabáj ‘apple’	*phabajénge/phabajénca*
púјka ‘turkey’	*pujkén*
žuv ‘louse’	*žuvénca*
noun with variation	
khanjí ‘hen’	*khanjén/khanján*

Table 48

Feminine nouns and their oblique forms from Vekerdi (1985)

The words from the newly collected data can be seen in Table 49. Words with no attested plural oblique form were excluded, and they are grouped together according to their oblique plural form, similarly to Table 48. Out of the total 20 items there are 4 whose oblique plural marker is *-an-*, there are 7 items whose oblique plural marker is *-en-*, and

there are 9 stems where the oblique forms vary. What is striking here is the fact that **the number of stems where there is variation is much higher** than expected based on Vekerdi (1985).

noun	attested oblique forms
nouns with the oblique plural *-an-*	
xajíng ‘well’	*xajingánge/xajingángo*
khajnʲí ‘hen’	*khajnʲán*
pīrí ‘saucepan’	*pīránge*
māčí ‘fly’	*māčánca*
nouns with the oblique plural *-en-*	
angrustí ‘ring’	*angrusténdar*
armajá ‘curse’	*armajénca*
cincə̃ri ‘mosquito’	*cincārénca*
kangrí/krangí ‘branch’	*kangrénca/krangénca*
kúrva ‘whore’	*kurvéngo*
mesají ‘table’	*mesajéndar*
rə̃ca ‘duck’	*rācén*
nouns with variation	
katt ‘a 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*
baj ‘sleeve’	*bājánca/bājénca*
bār ‘garden’	*bāránge/bārán/bārénge*
bórotva ‘razor’	*borotvénca/borotvánca*

Table 49

Feminine nouns and their oblique forms from the newly collected data

There is only a slight overlap between the two lists, but there are two, so all in all we have 9 lexical items with the oblique plural form *-an-* and 13 items with the oblique plural form *-en-*. The total number of stems where there is variation is 10. They can be seen in Table 50. The percentage shows the proportion of the two different forms among all the occurrences in the data (this chiefly means the newly collected data, as the data from Vekerdi 1985 contains only slight variation).

word	pl. obl. *-en-*	pl. obl. *-an-*
patrí ‘leaf’	70%	30%
papín ‘goose’	10%	90%
bāj ‘sleeve’	75%	25%
bār ‘garden’	57%	43%
šūrí ‘knife’	30%	70%
tʲīrí ‘ant’	45%	55%
katt ‘a pair of scissors’	22%	78%
māl ‘field’	25%	75%
borótva ‘borotva’	75%	25%
khanjí/khajnʲí ‘hen’	25%	75%

Table 50

The proportion of the two different forms among the occurrences in all the data at hand

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 look like this.

Figure 15

The proportion of the frequency of the stems belonging to the two feminine oblique plural patterns and the varying stems

We have to note here that Cech & Heinschink (1999) try to explain this again with the difference between inherited and borrowed words: *-an-* is used with inherited words and *-en-* is used for borrowed words. This is, however, completely inconsistent with the data and even with the way the inherited-borrowed dichotomy in the masculine is traditionally analysed, and thus should be dismissed.

The general frequency of /a/ and /e/ in the Romani verbal and nominal suffixes can play a role in 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 we could already see, while the vowels /u/ and /i/ appear less often in suffixes in general, and even then they are more typically used in derivation, /e/ and /a/ are quite common in the inflection of Romani, for example as the vowel component of nominal oblique markers, both feminine and masculine, and of personal concord markers on verbs.

As we can see in Table 51, the personal concord markers for consonantal verbs (with

the inclusion of the /e/ which was analyised as epenthetic) exclusively contain these two vowels.

	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*

Table 51

Verbal personal concord markers

If we consider the fact that the first and second person plural forms are less frequent generally, 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 feminine oblique plural.

The nominal oblique markers, including feminine nouns, can be *-es-*, *-en*, *-a-*, *-an-*, all containing /e/ or /a/. In addition, /o/ also appears in the variant oblique masculine forms *-os-* and *-on-*. The vowel /o/ is, however, not present elsewhere in the inflection. Considering all this, it follows that the variation in the feminine oblique plural between *-en-* and *-an-* is much more salient, with variation seen in 10 stems out of 32, than the variation in the masculine oblique between *-es-/-en-* and *-os-/-on*, where the number of stems where there is variation is only 9 out of 126.

It is also important to note that the variation always includes /e/ as one of the elements of 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. Let us not forget that it is always deleted where there is a thematic vowel at the end of the stem of the verb.

6.4.2 Possible causes and explanations

6.4.2.1 The masculine oblique plural -en-

The presence of the *-en-* pattern in the feminine may be connected to its simultaneous presence in the masculine. While the *-en-* pattern exerts a neutralising effect, making all plural paradigms look identical and decreasing the extent of gender difference, the *-an-* pattern exerts an opposite effect, trying to maintain an intra-gender uniformity, being more similar to the singular oblique marker *-a-*. A possible, additional aspect of variation is the presence of /n/ in the plural oblique across the whole nominal morphology; /n/ is a common trait of both the masculine and the feminine paradigms, so variation emerges more easily.

The influence of the masculine oblique plural *-en-* is shown in Figure 16, where the schemata for the masculine oblique plural and the feminine oblique plural are connected through a thick dashed arrow pointing from the masculine ending *-en-* to the feminine ending *-en-*.

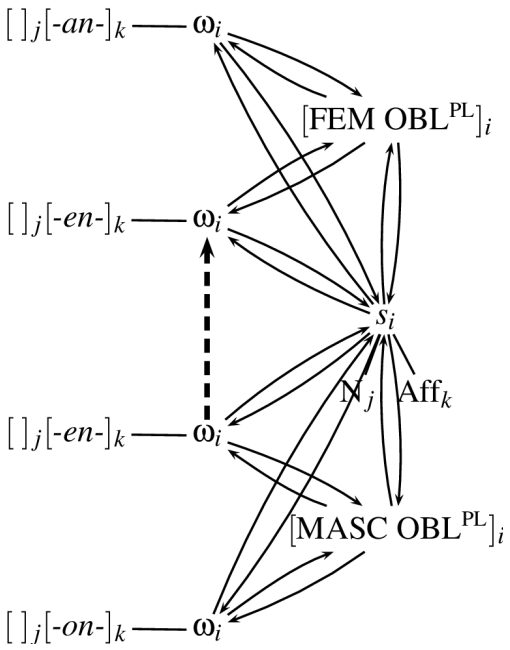

Figure 16

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’, which are apparently close cognates of each other, coming from Sanskrit *laḍikka* ‘child’, with the feminine form derived through gender assignment (see section 5.1.1).

nominative singular	nominative plural	oblique singular	oblique plural
rakló	*raklé*	*raklés-*	*raklén-*
rakljí	*rakljá*	*rakljá-*	*raklján-*

Table 52

Correlation between the masculine and feminine paradigms

As we can see from the example in Table 52, the 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 opposition of the nominative singular endings, /o/ and /i/, are swapped in the plural and in the oblique, but the front-back differentiation 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 take *-es-/-en-*. The high degree of the similarity of the two words in the nominative singular maintains the contrast, but in case the word *rakljí* had forms like **rakljén-* in the plural oblique, so if there were variation, it would not really be surprising to see forms such as **raklón-* for the word *rakló*.

Based on the newly collected data, the overall number of lexical items which have attested masculine oblique plural forms with the marker *-en-* is 40, as opposed to the 13 items with the marker *-on-*. Naturally, we can also add the masculine singular forms with the markers *-es-* and *-os-*, as they always correspond to each other, so the total number is 82 against 36 (not counting the stems where there is variation). If we compare this to the 13 feminine nouns with the oblique plural marker *-en-* and the 9 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, and the number of stems where there is variation is almost equal: 9 in the masculine and 10 in the feminine

paradigm. **The fact that there are more feminine nouns which take the *-en-* form suggests that the dominance of the *-en-* form in the masculine influences the feminine paradigm indeed.** The neutralisation effect is shown in Figure 17, where the masculine oblique plural and the feminine oblique plural converge in the ending *-en-*, and diverge through the endings *-an-* and *-on-*.

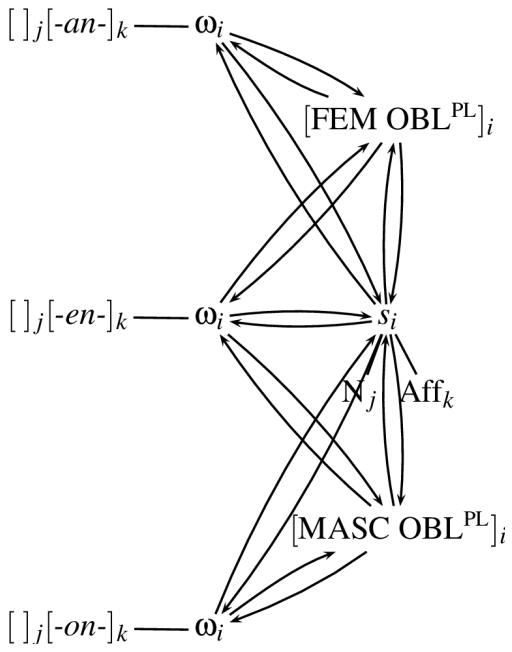

Figure 17

Combined schema of the masculine and feminine oblique plural

6.4.2.2 The feminine nominative plural suffixes

It would be appealing to say that the nature of the stem-final vowel plays a role in the choice of the oblique plural: if it is /i/, the vowel of the oblique plural marker is always /e/, if it is /a/, the vowel of the oblique plural marker is always /a/. However, as we could see from the data in Tables 48-50, this is definitely not the case. On the other hand, there might be a possible and even more obvious **correlation between the nominative plural and the oblique plural**. As we could see in Table 47, where the two patterns are introduced, the feminine plural form ends in /a/ if the nominative is /i/, so for example *pīrí* ‘pot, saucepan’ ~ *pīrá* ‘pots, saucepans’, and it ends in /i/ if the nominative is /a/, see *kúrva* ‘whore’ ~

kurví ‘whores’. The oblique forms seem to correspond to the plural forms as for their backness.[77]

(37)nominative singular *pīrí* → nominative plural *pīrá* → oblique plural *pīrán-*
nominative singular *kúrva* → nominative plural *kurví* → oblique plural *kurvén-*

If have a closer look at the data, we find the following numbers and proportions. Out of the total 33 items 16 items follow the pattern. This means that if the nominative plural ending is /i/, they will take the oblique plural ending *-en-*, and if the nominative plural ending is /a/, they will take the oblique plural ending *-an-*, as seen in Table 53.

noun	nominative plural form	oblique plural form
nouns with the oblique form *-an-*		
phenʲ ‘sister’	*phenʲá*	*phenʲán-*
raklʲí ‘girl’	*raklʲá*	*raklʲán-*
rolʲí ‘stick’	*rolʲá*	*rolʲán-*
romnʲí ‘woman’	*romnʲá*	*romnʲán-*
šej ‘girl’	*šejá*	*šején-*
žuvljí ‘woman’	*žuvljá*	*žuvlján-*
xajíng ‘well’	*xajingá*	*xajingán-*
māčí ‘fly’	*māčá*	*māčán-*
pīrí ‘saucepan’	*pīrá*	*pīrán-*
nouns with the oblique form *-en-*		
cóxa ‘skirt’	*coxí*	*coxén-*
pápuča ‘shoe’	*papučí*	*papučén-*
púјka ‘turkey’	*pujkí*	*pujkén-*
armajá ‘curse’	*armají*	*armajén-*
kúrva ‘whore’	*kurví*	*kurvén-*
rə̄ca ‘duck’	*rācí*	*rācén-*

Table 53

Feminine nouns where the nominative plural ending corresponds to the oblique plural ending

77 Although throughout its history, Romani has been in contact with several languages with vowel harmony, like Turkish and Hungarian, there is no reason to suppose that the backness correspondence here is more than mere coincidence.

9 items behave in the opposite way, so their nominative plural ending is /a/ alongside the oblique plural ending *-en-*. There are no nouns whose nominative plural ending would be /i/ alongside the oblique plural ending *-an-*.

noun	nominative plural form	oblique plural form
birulʲí ‘bee’	*birulʲá*	*birulʲén-*
čerháj ‘star’	*čerhajá*	*čerhajén-*
čiriklí ‘bird’	*čiriklʲá*	*čiriklʲén-*
phabáj ‘apple’	*phabajá*	*phabajén-*
žuv ‘louse’	*žuvá*	*žuvén-*
cincā̆ri ‘mosquito’	*cincārá*	*cincārén-*
mesají ‘table’	*mesajá*	*mesajén-*
angrustí ‘ring’	*angrustá*	*angrustén-*
kangrí/krangí ‘branch’	*kangrá/krangá*	*kangrén-/krangén-*

Table 54

Feminine nouns where the nominative plural ending does not correspond to the oblique plural ending

The difference is significant, with almost twice as many items where there is correspondence in the backness.

Let us also check the tendencies among the 10 stems where there is variation. 7 of the stems where there is variation predominantly take either the nominative plural ending /a/ and the oblique plural ending *-an-*, or the nominative plural ending /i/ and the oblique plural ending *-en-*.

word	pl. obl. *-en-*	pl. obl. *-an-*
papín ‘goose’	10%	90%
šūrí ‘knife’	30%	70%
tʲīrí ‘ant’	45%	55%
katt ‘a pair of scissors’	22%	78%
māl ‘field’	25%	75%
borótva ‘borotva’	75%	25%
khanjí/khajnʲí ‘hen’	25%	75%

Table 55

Feminine nouns where there is variation with a bias towards the correspondence between the nominative plural and the oblique plural in backness

On the other hand, only 3 of the stems with varying forms go against the tendency, with the predominant pattern being that of the combination of the nominative plural ending /a/ and the oblique plural ending *-en-*.

word	pl. obl. *-en-*	pl. obl. *-an-*
patrí ‘leaf’	70%	30%
bāj ‘sleeve’	75%	25%
bār ‘garden’	57%	43%

Table 56

Feminine nouns where there is variation with no bias towards the correspondence between the nominative plural and the oblique plural in backness

In total, we can say that the nominative plural ending can definitely or predominantly predict the corresponding oblique plural for 23 stems, while this prediction goes awry in case of only 10 stems. This suggests that **there is a tendency for the feminine nominal plural suffix to influence the choice of the oblique plural suffix**, but it might be weakened by the fact that the nominative singular suffix is exactly the other way round. This is shown in Figure 18, where the schemata 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.

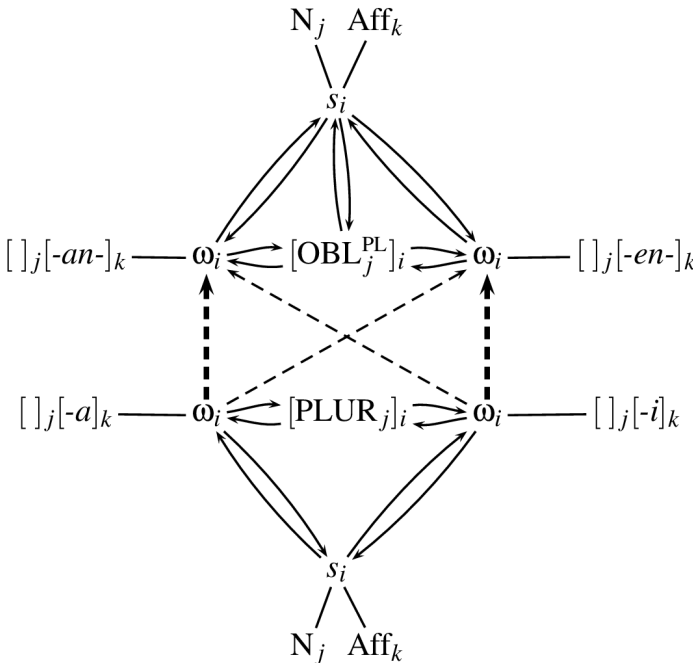

Figure 18

The relationship between the feminine nominative plural and the feminine oblique plural as shown in the form of schemata

6.4.3 Conclusion

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 seem to exert influence indeed.

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 the feminine one. 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-*. We also found an overall dominance of the marker *-en-*.

6.5 Weak points in the Lovari verbal paradigms

In this section, we will first look at the present paradigms and the possible analogical effects. Then we will discuss the third weak point, the past tense of vocalic verbs, that is, verbs with either a stem-final /a/ or /i/ in detail.

6.5.1 Some more remarks on the present tense

We must mention that, although the paradigm of *-i-* stem verbs as shown in section 5.2.2 has been confirmed by the newly collected data, some sources (e.g. Cech & Heinschink 1999 and Boretzky 1994) claim that "longer" or "full" forms exist simultaneously in other Vlax Romani varieties.[78] If we take a closer look at Table 57, which presents both the "full" and the contracted forms, we can see two things right away. Firstly, the "full" forms are very similar to the consonantal verbs, insofar as their stem ends in a consonant, and as long as the personal concord markers are considered. The elements or markers (*-(V)sar* and *-(V)n*, respectively) that appear before the personal concord markers will also be touched upon in section 6.5.2.1. Based on the newly collected data, though, **there is no variation in the present tense in Hungarian Lovari**. Secondly, the Kalderash form *gîndív* in the first person singular suggests that the consonant featured in the personal concord markers of the most numerous consonantal class and the *-a-* stem class can spread onto the verbs of

78 This is in line with the explanation Matras (2002) gives about its origins.

the *-i-* stem class, too.

Cech & Heinschink (1999), Hutterer & Mészáros (1967) and Matras (2002) all consider the verbs with the thematic vowel /i/ as loan verbs (mostly from Romanian and Hungarian), the paradigms of which, differing both from that of the consonantal class and that of the vocalic class, was created through contraction. If we seriously consider, just for a moment, the possibility of contraction, we face several problems. It is hard to explain how the form *gindináv* became *gindíj* (as stated in Cech & Heinschink (1999) and contrary to the fact that the form *gîndisaráv* became *gîndív*, as claimed by Boretzky 1994), and why the form *gindinás* was kept in the first person plural while the form *gîndisarás* became *gîndís* in Kalderash. The description is ambiguous, too: Cech & Heinschink (1999) say that in Austrian Lovari, the paradigm of the "full" forms is more common, while the contracted forms are typical of Kalderash and seldom appear in Austrian Lovari.

present tense	(Austrian) Lovari		(Serbian) Kalderash	
indicative	"full" forms	contracted forms	"full" forms	contracted forms
singular	gindináv	gindíj	gîndisaráv	gîndív
	gindinés	gindís	gîndisarés	gîndís
	gindinél	gindíj	gîndisarél	gîndíl
plural	gindinás	gindinás	gîndisarás	gîndís
	gindinén	gindín	gîndisarén	gîndín
	gindinén	gindín	gîndisarén	gîndín

Table 57

The "full" forms and the contracted forms of *-i-* stem verbs

There is no doubt that these verbs (for example *gindí-* 'think', *vorbí-* 'speak', *trají-* 'live') are originally loanwords. But we have to bear in mind that the native speaker has no access whatsoever to any etymological information. What they have at hand are patterns, some of which are perhaps more dominant than the others. It is very natural for languages to employ a derivational marker for inserting loanwords. However, this does not seem to be the case with Lovari, or at least only partly. Cech & Heinschink (1999) list several loan verbs in the form of an *-i-* stem verb, like H. *kereskedik* > *kereškedí-* 'trade'; H. *pihen* >

pihení- ‘rest’; H. *átkoz* > *ātkozí-* ‘curse’.

These come in the form of regular *-i-* stem verbs, rather than inromani tenses a form containing any derivational marker (*-(V)n-* or *-(V)sar*). The newly collected data confirm this, as a fair number of verbs, of various origins, appear with a stem-final /i/.

1st person singular:

(38)*kaná* *khēr* *aváv* ***cipíj***
‘when’ ‘home’ ADV. ‘come’ 1st SING. PRES. IND. ‘shout’ 1st SING. PRES. IND.
‘When I arrive home, I will shout.’

2nd person singular:

(39)***imātkozís*** *andá* *mandé*
‘pray’ 2nd SING. PRES. IND. ‘for’ PREP. 1st SING. PERS. PRON. LOC.
murí *dej?*
1st SING POSS. PRON. ‘mother’ NOM. SING.
‘Will you pray for me, mother?’

(40)*soskó* *dezodorí* ***hasnālís***
‘what kind of’ ‘deodorant’ NOM. SING. ‘use’ 2nd SING. PRES. IND.
‘What kind of deodorant do you use?’

3rd person singular:

(41)*o* *rašáj* ***āldíj***
ART. DEF. M. SING. ‘priest’ NOM. SING. ‘bless’ 1st SING. PRES. IND.
e *ternén*
ART. DEF. PL. ‘young’ ADJ. PL.
‘The priest blesses the young ones.’

2nd person plural:

(42)*so* ***njerín*** *tumé* *te*
'what' INT. 'win' 2nd PL. PRES. IND. 2nd PL. PERS. PRON. NOM. 'if' CONJ.
njerína *po (pe o)* *veršenjí*
'win' 2nd PL. FUT. IND. 'on' PREP. + ART. DEF. 'competition' NOM. SING.
'What will you win if you win the competition?'

3rd person plural:

(43)***kēšín*** *e* *gēzéša*
'be late' 3rd PL. PRES. IND. ART. DEF. PL. 'train' NOM. PL.
'The trains are [always] late.'

Some more examples:

(44)R. *ruga* > *ruďí-* 'pray'
R. *scrie* > *iskirí-* 'write'
R. *ajuta* > *žutí-* 'help'
H. *tetszik* > *tecí-* 'please, be liked'
H. *ápol* > *āpolí-* 'tend, care'
H. *mobilozik* > *mobilozí-* 'use/play with one's mobile phone'
H. ?*ásózik* > *āšōzí-* 'dig'
H. *bír* > *birí-* 'can, be able to'
H. *használ* > *hasní-/hasnālí-* 'use'
H. *késik* > *kēší-* 'be late'
H. *rúg* > *rugí-* 'kick'
H. *bízik* > *bizí-* 'trust'
H. *emlékszik* > *emlēksí-* 'remember'
H. *kezdődik* > *kezdēdí-* 'begin' INTR.
H. *utál* > *utālí-* 'hate'
H. *tagad* > *tagadí-* 'deny'

In Hungarian Lovari, the *-i-* stem class is the direct landing site of new borrowings, but at the same time it contains many items whose origins are not obvious for a native speaker who is only in contact with Hungarian. This leads us to claim that the *-i-* stem verbs have indeed come to form a verb class in their own right, as stated in section 5.2.2.

More evidence that variation occurs where there are weak points in the grammar comes from the word meaning 'save'. Perhaps due to the influence of the Hungarian word *ment* meaning the same thing, it appears in several forms in the newly collected data.

(45)*e* *doktorá* *trajó* *mentón/mentín/muntún*
ART. DEF. PL. 'doctor' NOM. PL. 'life' NOM. SING. 'save' 3rd PL. PRES. IND.
'Doctors save lives.'

muntún	*mentín*	*mentón*
10%	50%	40%

Table 58
The proportion of the different forms *mentí-/mentó-/muntú-* in the newly collected data

This variety of forms does not only reflect lexical variation. On the one hand, it shows that low token frequency is a trigger of variation; on the other, it marks the obvious dominance of the *-i-* stem verbs.

6.5.2 Variation in the past tense of vocalic verbs

In this section, we will discuss the variation seen in the past tense of verbs with a stem-final /a/ or /i/. After the description of the phenomenon, we will examine two possible aspects that might influence their past forms. The two aspects are the following.

1. The stem-final /r/, /n/ and /v/ of consonantal verbs. Additional sounds or sound sequences, which also resemble certain derivational markers, appear in the past forms of

vocalic verbs besides one of the regular past tense markers. We will explain their appearance with the lack of a straightforward pattern on the one hand, and the analogical effect of certain consonantal verbs on the other, and we will find that the relationship, at least in terms of token frequency, can clearly be seen.

2. The past forms of verbs with the derivational markers *-av-* and *-ajv-*. We will look at the possible connection between the past forms of verbs with the derivational markers *-av-* and *-ajv-* and the past forms of vocalic verbs. The results are convincing for the *-i-* stem verbs, but not so convincing for the *-a-* stem verbs, so this will definitely need to be reconsidered in the future.

6.5.2.1 Description of the phenomenon

As we could see in section 5.2.3, the formation of the past tense of the consonantal class is very regular and consistent. This is represented by the following schema, shown in Figure 19, where V denotes a verb, C_A refers to the set of consonants containing the voiced alveolars, while consonant set C_B contains the bilabials, velars and voiceless fricatives. We can see from the schema in Figure 19 that there is correspondence between the two past tense markers and the fact that there is a consonant stem-finally (consonant set C_A exclusively takes the marker *-d-*, while consonant set C_B exclusively takes the marker *-l-*), but nothing is said about verbs with a stem-final vowel.

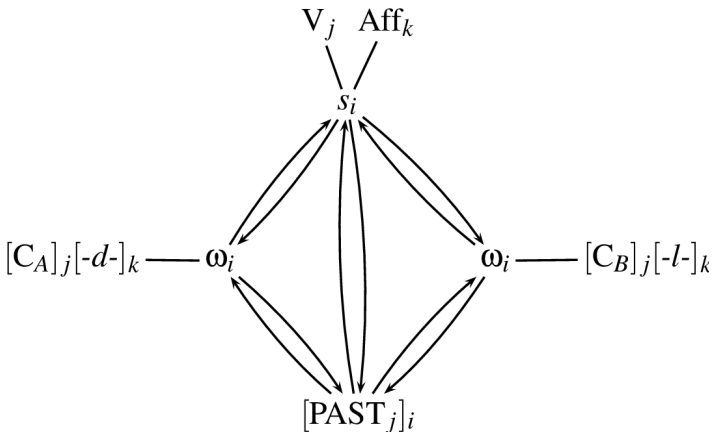

Figure 19

Schema for the past tense of the consonantal class

As it can be seen from the schema, t**here is a clear-cut correspondence between the phonological aspect that there is a consonant at the end of the stem** and the function or semantic content of the past tense. **However, the lack of a stem-final consonant and the presence of a vowel at the end of the stem causes disturbance in the formation of the past tense**, and the unambiguous correspondence between the past function and the phonological form of the markers becomes weaker.

We could consider the past tense of the vocalic classes separate schemata, but the degree of the variation shown by them in the past tense, along with the few verbs exhibiting a stem-final /o/ and /u/ is so high and the variation itself is so unpredictable that the correspondences would be too weak. Rather, we regard this as an act of pattern seeking (cf. Blevins & Blevins 2009), where the past tense schema will also serve as one of the patterns.

The variation in the -*a*- stem class is shown in Table 59. The three different paradigms appear interchangeably within Lovari.

past tense indicative	*asá*- ‘laugh’		
singular	asajém	asadém	asandém
	asaján	asadán	asandán
	asajás	asadás	asandás
plural	asajám	asadám	asandám
	asaján	asadán	asandán
	asajé	asadé	asandé

Table 59

The possible past paradigms of -a- stem verbs

The following 18 -*a*- stem verbs in (46) are listed in Vekerdi (2000).

(46)*asá*- ‘laugh’

azbá- ‘hurt’

bilá- ‘melt’

dará- ‘be afraid’

dukhá- ‘ache’

fimlá- ‘glitter’

xa- ‘eat’

xasá- ‘cough’

izdrá- ‘tremble’

ladžá- (pe) ‘be ashamed’

langá- ‘limp’

lošá- ‘be glad’

patʲá- ‘believe’

pitʲá- ‘drip’

prasá- ‘mock’

sungá- ‘smell’

tromá- ‘dare’

urá- ‘fly’

ža- ‘go’

The newly collected data yielded the results shown in Table 60, based on the verbs scrutinised in the questionnaire.

verb	*-j(l)-*	*-d-*	*-nd-*
asá- ‘laugh’	33%	44%	23%
lošá- ‘be glad’	73%	0%	27%
tromá- ‘dare’	70%	0%	30%
prasá- ‘mock’	50%	25%	25%

Table 60

The proportion of the different past tense forms of the *-a-* stem verbs in the newly collected data

Four additional verbs, *azbá-* ‘hurt’, *langá-* ‘limp’, *xasá-* ‘cough’ and *patʲá-* ‘believe’ were included in the questionnaire, too. The information obtained about *langá-* ‘limp’, *xasá-* ‘cough’ was not sufficient to draw any conclusions. As for the other two verbs we found no variation: *azbá-* ‘hurt’ only has past forms in *-d-*, while *patʲá-* ‘believe’ only has

past forms in *-j(l)-*. The verb *žá-* is excluded due to the suppletive nature of its past form (*gel-*), but it basically belongs to the group of verbs with a perfective marker *-l-*. In addition to that, Vekerdi (1985) contains some more past tense forms, such as *xal-*, *ladžal-*, *sungal-*, *ural-* and *dukhal-*, but it is not completely clear whether these are the perfective stems or the conditional form, which corresponds to an imperfective aspect. The verb *dará-* 'be afraid' is attested in Vekerdi (1985) with the past form *darajl-*. The other verbs were not attested. An additional verb, *colaxá-* 'get married' occurred in the newly collected data. The documented form of this verb, however, is *colaxar-*, so this needs further investigation.

Thus, the overall proportion of the frequency of the different perfective markers for *-a-* stem verbs looks like this.

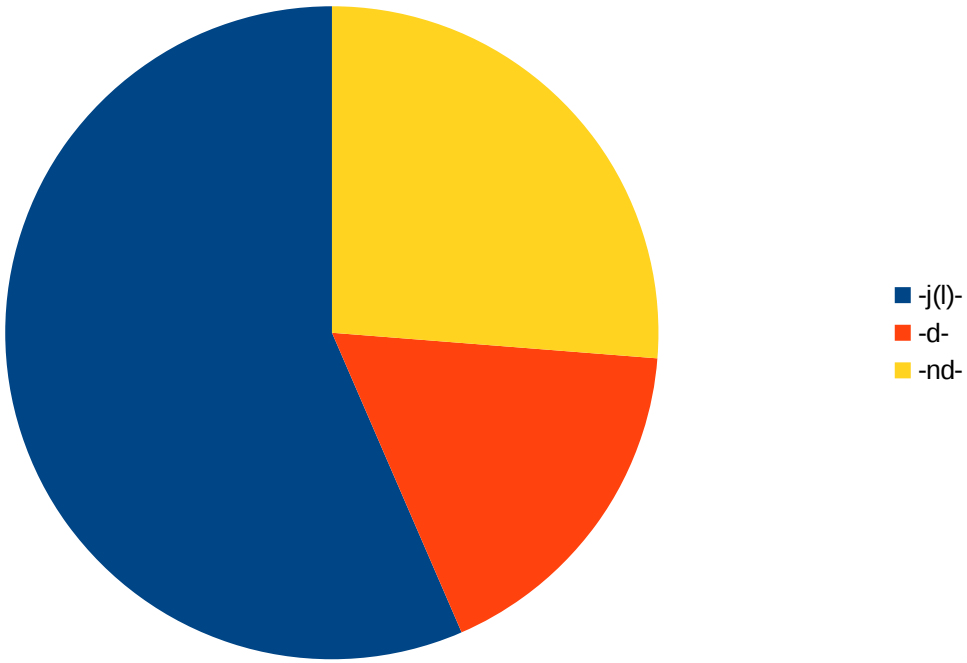

Figure 20

The proportion of the frequency of the different perfective markers for *-a-* stem verbs

There is significant variation in the past tense of *-i-* stem verbs, too, but only two dominant patterns compete here, and an additional, minor pattern, as opposed to the three patterns we see in case of the *-a-* stem verbs. The inflection according to the three patterns can be seen in Table 61 (the third column is the minor, but existing pattern).

past tense indicative	*gindí-* ‘think’		
singular	gindisardém	gindindém	gindisajlém
	gindisardán	gindindán	gindisajlán
	gindisardás	gindindás	gindisajlás
plural	gindisardám	gindindám	gindisajlám
	gindisardán	gindindán	gindisajlán
	gindisardé	gindindé	gindisajlé

Table 61

The possible past paradigms of the *-i-* stem verbs

The verbs under scrutiny can be seen in Table 62.

verb	*-nd-*	*-sard-*	*-sajl-*
žutí- ‘help’	20%	80%	0%
cipí- ‘shout’	25%	75%	0%
iskirí- ‘write’	36%	64%	0%
igērí- ‘promise’	25%	75%	0%
rugí- ‘kick’	22%	78%	0%
rudjí- ‘pray’	20%	80%	0%
āldí- ‘bless’	34%	66%	0%
birí- ‘be able to’	83%	17%	0%
gindí- ‘think’	12%	55%	33%
tecí- ‘please, appeal, be liked’	25%	25%	50%
buji- ‘hide’	0%	80%	20%

Table 62

The proportion of the different past tense forms of the *-i-* stem verbs in the newly collected data

The overall proportion of the frequency of the different perfective markers for *-i-* stem verbs looks like this.

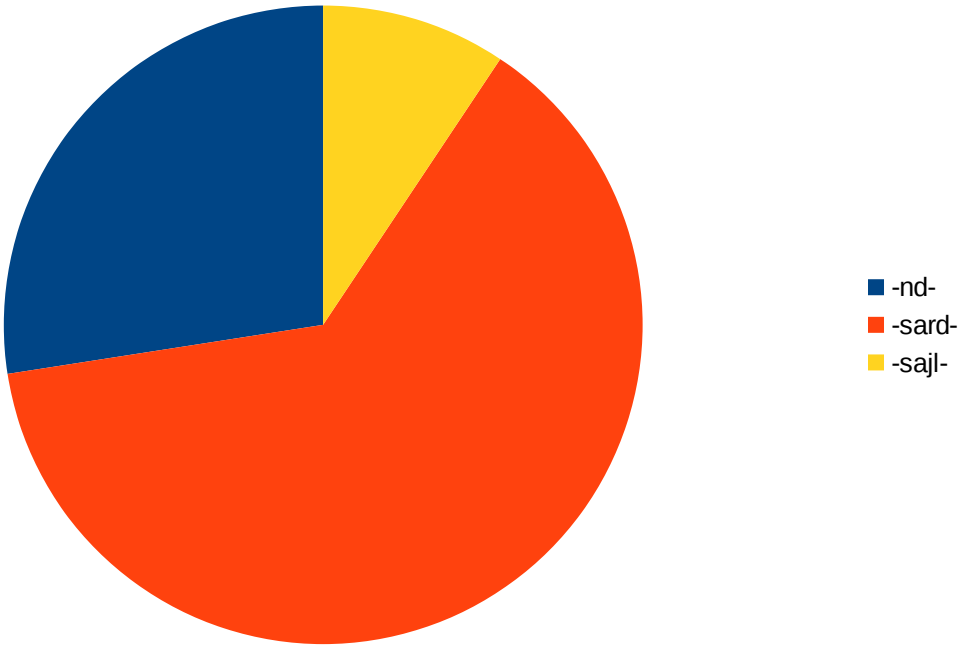

Figure 21

The proportion of the frequency of the different perfective markers for *-i-* stem verbs

Other, sporadic examples include the past indicative forms in (47).

(47) *n^{j}erisardé* ‘win’ 3rd plural
senvedisardém ‘suffer’ 1st singular
mentisardé ‘save’ 3rd plural
ātkozisardás ‘curse’ 3rd singular
imātkozisardán ‘pray’ 2nd singular/plural
zūditindás ‘overwhelm’ 3rd singular
māsindám ‘climb’ 1st plural
akastisardém ~ *akastindém* ‘hang’ 1st singular

kapisardás ~ kapindás ‘touch’ 3rd singular
lekerelisardás ~ kerülisardás [angle] ~ kerülindás [angle] ‘turn up’ 3rd singular
skepisajlás ‘escape’ 3rd singular

We should note in passing that several attempts have been made to formalise the past tense of these verbs. Hutterer & Mészáros (1967) give forms similar to the past tense of mediopassive verbs: *patʲá-* → *patʲil-*. But as this is the only example given, we can hardly be sure that this is a general pattern. Moreover, this pattern does not appear anywhere in the newly collected data. The pattern might have been heard, as the verb given as an example contains a palatal before the thematic vowel, which could have triggered forms like that in the speech of a speaker who did not have a native level command of the language. However, the forms that appear in the newly collected data more or less correspond to the data we find in Cech & Heinschink (1999) and Boretzky (1994). Matras (2002), while providing a fairly detailed overview of the possible markers of ‘psych verbs in *-a*’ (Matras 2002: 142) in Romani in general (*-il-/-jl-*, *-in-/-n-*, *-d-* and even the combinations of these), does not give any explanation for the variation we see in Lovari (and probably in other dialects, too).

In case of the consonantal class, the consonants provide a very clear basis for the choice of the perfective marker. **In case of the vocalic classes**, there is no guiding consonant, so the paradigms become unstable and **several patterns enter into competition**. Three patterns can be seen here. The first one, *-j(l)-*, corresponds to the perfective marker *-l-*, and also to the pattern followed by certain stems ending in /v/, especially the one with the marker *-ajv-*. The influence (marked with arrows) here must be enhanced by the /a/ appearing in the marker, which is the thematic vowel of the class in question.

perfective *xojajl-* ‘get angry’ → perfective *tromájl-* ‘dare’ ← perfective *maladʲil-* ‘meet’
↑
perfective *ušťʲil-* ‘get up’

Figure 22
Influences among the past tense forms

The second one, *-d-*, is the marker of consonantal stems ending in a voiced alveolar. An additional analogical effect here is the fact that the marker *-d-* frequently follows a vowel, as it triggers the deletion of the stem-final /v/ and /d/. **The frequency effect is further increased** by the numerous verbs derived with the help of the transitive marker *-av-*, whose vowel /a/ is identical to the thematic vowel of the *-a-* stem class, and whose perfective stem thus ends in the sequence *-ad-*.

bašav- 'play music' ~ perfective *bašad-* → *asá-* 'laugh' ~ perfective *asad-*

Figure 23
More influences among the past tense forms

The third pattern contains a surprising additional consonant, /n/. It is interesting to note its similarity to the marker *-in-*. As we will see in section 6.5.2.2, /n/ appears in the perfective forms of the other vocalic class, too, following the stem-final /i/, which results in an interaction between the two vocalic classes.

perfective *vorbind-* 'speak' ↔ perfective *tromand-* 'dare'

Figure 24
Influences among the *-i-* and *-a-* stem verbs

The verbs with only one attested paradigm (*azbá-* 'hurt' and *patʲá-* 'believe') are also worth noting, as the fact that they only inflect in one way might be the result of a high token frequency which makes one paradigm strong enough not to trigger variation.

The reasons for the variation in the *-i-* stem class are diverse. In spite of the fact that the *-i-* stem class is in constant motion because it is the landing site of newly coined borrowings, due to the large number of constantly changing verbs and due to some very frequent verbs, so due to both high token and type frequency, it seems to constitute a fairly uniform and solid class. Another reason is that there are less analogical effects of the phonological sort than for the *-a-* stem verbs (see the discussion on the marker *-av-* in

section 5.2.4.1).

Variation here is not limited to the dialect, it exists within one speaker, too; as if the speaker had been looking for the “right” form, correcting themselves, but ultimately could not find it because both of them are equally “correct”.

(48)*gindisardem* *hodj* *rumuj* *pe*
‘think’ 1st SING. PAST IND. ‘that’ CONJ. ‘go wrong’ 3rd SING. PRES. IND. PRON. REFL.

(49)*gindisajlem* *hodj* *rumuj* *pe*
‘think’ 1st SING. PAST IND. ‘that’ CONJ. ‘go wrong’ 3rd SING. PRES. IND. PRON. REFL.

Two patterns contain derivational markers presented in section 5.2.4, one pattern shows the consonant /n/, seen in Table 59 among the past forms of *-a-* stem verbs, also possibly and originally some sort of a marker.

The marker *-in-* is indeed used to adapt loan verbs by placing them in the consonantal class in other varieties of Romani spoken in the region, like Romungro, which belongs to the Central dialect group.[79]

(50)*me* *lil*
1st PERS. SING. PERS. PRON. NOM. ‘letter’ NOM. SING.
īrináv
‘write’ 1st PERS. SING. PRES. IND.
‘I write a letter.’

79 We could also say that loan verbs are inserted into these varieties of Romani with the help of a loan marker, which would come as no surprise, as similar phenomena happened before, cf. Bakker (1997). However, historical linguistics becomes obscure here and therefore does not necessarily support this assumption. The exact source of the marker *-in-* used in this function is not clear; it can go back to the Greek present tense markers but also to the Old Indo-Aryan participle marker. Some sources mention that it appears in Hungarian Lovari as an independent loan-verb adaptation marker (e.g. **boksolin-* ‘box v.’ from Hungarian *bokszol*, **birkozin-* ‘wrestle’ from Hungarian *birkózik*, with the addition of the marker to the bare stem *birkóz-*). The marker could have been borrowed from neighbouring Central dialects into Lovari, but fresh data do not confirm its presence as a loan-verb adaptation marker. Loan-verbs almost exclusively land in the *-i-* stem class.

(51)*mindēg* *ōvatošan*

‘always’ [direct borrowing from Hun.] ‘carefully’ [direct borrowing from Hun.]

trad *kaná* ***vezetinés***

‘drive’ 2nd PERS. SING. IMP. ‘when’ PRON. REL. ‘drive’ 2ND PERS. SING. PRES. IND.[80]

As shown in Table 57, it also appears as part of an alternative paradigm in Austrian Lovari, but no similar variation is present in the present paradigms of Hungarian Lovari – neither with the forms containing the marker *-in-*, nor with the forms containing *-sar-* (which vary in Kalderaš). Historically, the markers might have been present in the present paradigms, too; this might be preserved by its presence in the 1st person plural. Now we can only say that the 1st person plural of *-i-* stem verbs with its different form maintains a contrast with the 2nd person singular. Its presence there, however, could be a basis for its appearance in the past paradigm.[81] So it is not the forms containing the marker *-sar-* that become contracted; it is the other way round: **the *-i-* stem *v*erbs with high type frequency**, albeit with a constantly changing membership, on top of the members with high token frequency (*vorbí-* ‘speak’, *gindí-* ‘think’ etc.) **attract all the new verbs**. In the past tense, however, where there is no one single, sufficiently solid pattern, only the consonantal class, **it is exactly the consonantal class with high type frequency that will begin to attract all other verbs**, and this is how the forms with the markers are created, making them similar to the consonant-final verbs. This does not only make it easier to explain why the first person singular form is typically *gindíj* etc. – it is a separate class, which differs from the consonantal and the *-a-* stem verbs in this respect, while having its phonological reasons, as discussed in section 5.2.2.1 –, but also why *gindív* can emerge in Kalderaš, which never seems to occur in Lovari).

Although it was not in the scope of the questionnaire, some data arose concerning the

80 The data were collected in Hungary by the author for the Linguistic Atlas of Central Romani under the auspices of Charles University in Prague.

81 It must be noted that the imperative of these verbs also contains the marker *-sar-* regularly: *na kodó* ***gindisár*** *hodʲ*… ‘don’t think that…’; ***kezdisár*** *te kheles tu* ‘start to play!’; ***bočājtisár*** *mande* ‘forgive me!’ etc. It is obviously the same in Romungro: ***vidʲāzín*** ‘take care!’; ***fejezín*** *mā andé adaná dilinipjá* ‘finish those silly things at last!’. Most probably because of the low frequency of the category again, there seems to be variation even within Lovari. In the collected Lovari data, there are examples which correspond to the perfective marker *-nd-*: ***bočātín*** *amenge* ‘forgive us!’. The fact comes as no surprise, as imperative forms with a final vowel are very rare (the imperative of most consonantal verbs is the bare stem, except verbs with a stem-final palatal or affricate – cf. the Romungro examples well-known from Hungarian as well: *ušťʲi* ‘get up!’ and *āčʰi* ‘stop!’).

past tense of the small number words with an apparent stem-final /u/, which suggest that rare items pair up with rare patterns: *rumusajlás* ‘go wrong’ 3rd person singular past indicative and *bunusajlém* ‘regret’ 1st person singular past indicative. The possible past paradigm would look like this.

past tense indicative	*bunú-* ‘regret’
singular	bunusajlém bunusajlán bunusajlás
plural	bunusajlám bunusajlán bunusajlé

Table 63

The past paradigm of the rare verbs with a stem-final /u/

The verbs with a stem-final /u/ also serve as a good argument for the influence of both the *-i-* stem verbs and the consonantal class. The -u- stem class, similarly to the -i- stem class, is exclusively made up of loan verbs, but it is very small. Apart from the modal auxiliary *trubú-* ‘must, need’ and the main verb *rumú-* ‘go wrong’, which exist for every speaker in the newly collected data, these verbs either do not exist or are replaced by other forms (cf. the example *muntú-* in section 6.5.1). The present tense paradigm of these verbs show similarity to the vocalic classes, and the past tense paradigms, too, but this also means that they follow the pattern of the consonantal verbs. On the other hand, the complete lack of variation both in the present and in the past tense suggests that the existing verbs represent a small but solid group, perhaps owing to the nature of the modal auxiliary.

If we now take a glance at the various possible forms of *-i-* stem verbs in Hungarian Lovari, we find the following, shown in Table 64, through the word *tecí-* ‘please, appeal, be liked’. It seems that the present tense forms can go with any of the past paradigms, and even the same speaker might mix the different paradigms.

indicative	present tense	past tense		
singular	tecíj	tecindém	tecisajlém	tecisardém
	tecís	tecindán	tecisajlán	tecisardán
	tecíl	tecindás	tecisajlás	tecisardás
plural	tecisarás	tecindám	tecisajlám	tecisardám
	tecín	tecindán	tecisajlán	tecisardán
	tecín	tecindé	tecisajlé	tecisardé

Table 64

Possible past paradigms of the -*i*- stem verbs

All in all, one thing is clear. **There are at least two different kinds of analogical forces "competing" in the verbal system**: that of the consonantal class and that of the -*i*- stem class, with additional influences from certain derivational markers.

6.5.2.2 Possible causes and explanations

6.5.2.2.1 -*i*- The stem-final /r/, /n/ and /v/ of consonantal verbs

Due to the presence of the same consonants, there is a possibility that the stem-final /r/, /n/ and /v/ of consonantal verbs influences the past forms of vocalic verbs. This applies to the -*a*- stem verbs on the one hand, where a perfective marker in the form of -*nd*- appears, as well as to the -*i*- stem verbs, where a past form which would traditionally be called regular does not exist at all. All existing past forms are formed with the addition or insertion of extra elements, consonants or strings, which resemble derivational markers. In this respect, the question where the elements come from is not relevant. The language user will turn to an existing inventory which contains these items. Again, the fact that they also function as derivational markers just means that the verbal system has to be viewed more holistically, as a more complex and interrelated set of elements. The reason why these particular items are chosen could be found in surface similarity. **The appearance and use of the markers -*sar*- and -*(V)n*- can be related to the consonantal verbs with a stem-final /r/ or /n/.** As we mentioned already, the pattern we see in the overwhelming class of consonantal verbs does not provide enough clue as to what to do with verbs whose stem ends in a vowel. Thus, it is no surprise that there is uncertainty about the forms.

Consonantal verbs with a stem-final /v/ might also influence, even if only to a lesser extent, the past forms of *-a-* stem verbs. As mentioned in section 5.2.3, the stem-final /v/ is deleted, giving rise to perfective stems which end in the sequence *-(V)d-*, and that is exactly what we see in one variant of the *-a-* stem verbs. Verbs like that include *thov-* 'wash', *paruv-* 'change'.

	present stem	past stem
consonantal verb with a stem-final /r/	*mar-* 'beat'	*mard-*
-i- stem verb	*žutí-* 'help'	*žutisard-*
consonantal verb with a stem-final /n/	*tʲin-* 'buy'	*tʲind-*
-i- stem verb	*žutí-* 'help'	*žutind-*
consonantal verb with a stem-final /n/	*an-* 'bring'	*and-*
-a- stem verb	*tromá-* 'dare'	*tromand-*
consonantal verb with a stem-final /v/	*thov-* 'wash'	*thod-*
-a- stem verb	*prasá-* 'mock'	*prasad-*

Table 65

The vocalic verbs and their possible consonantal influences

The effects presented in Table 65 can also be shown in the form of schemata. Figure 25 shows the influence of the stem-final /r/ on the presence of the past tense marker *-sard-*, Figure 26 shows the relationship between consonantal verbs with a stem-final /v/ and the appearance of the past tense marker *-d-* on *-a-* stem verbs, while Figure 27 shows the possible influence of consonantal verbs with a stem-final /n/ on the past tense marker *-nd-*.

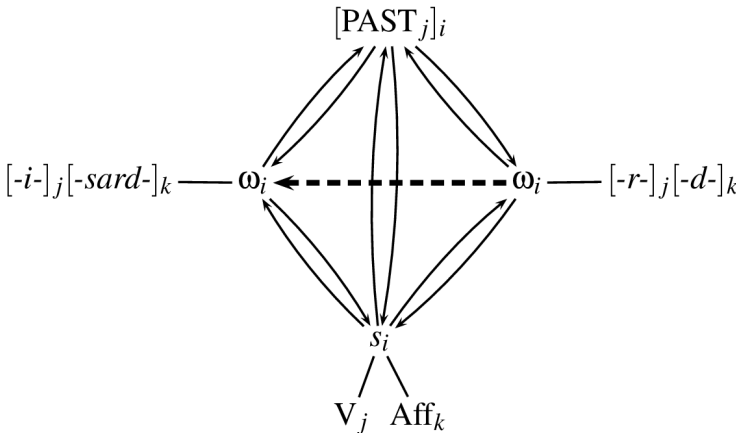

Figure 25

The stem-final /r/ and the marker -*sard*-

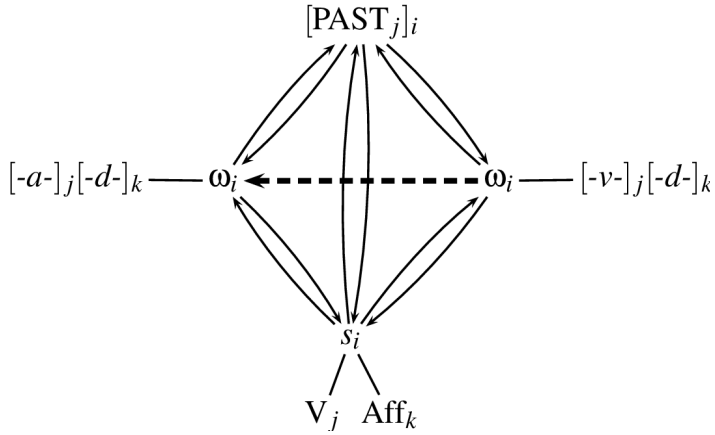

Figure 26

The stem-final /v/ and the marker -*d*- on -*a*- stem verbs

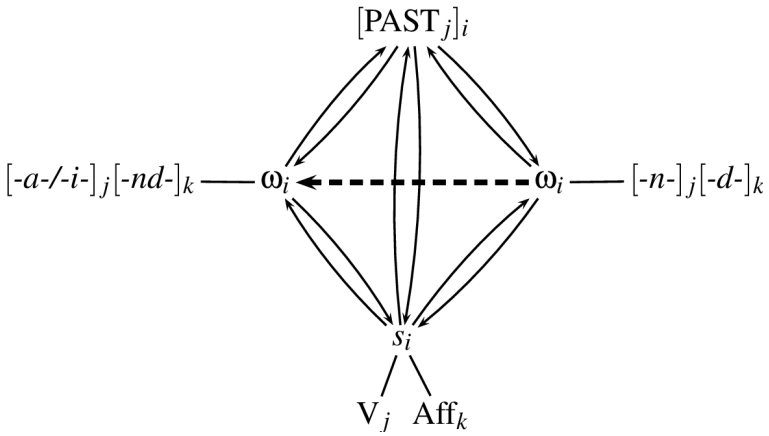

Figure 27

The stem-final /n/ and the marker *-nd-*

There are 32 consonantal verbs with a stem-final /r/, 9 with a stem-final /n/ and 6 with a stem-final /v/, not counting the verbs with the derivational markers *-av-* and *-ajv-* (for details about that, see section 6.5.2.2.2).

Let us take first the proportion of consonantal verbs with a stem-final /r/ and a stem-final /n/, which is 3.5:1. In comparison, we will have a look at two things here: the overall number of *-i-* stem verbs with attested past forms taking the perfective markers *-sard-* and *-nd-* in the newly collected data and in Vekerdi (1985) on the one hand, and the proportion of the occurrences of the perfective markers *-sard-* and *-nd-* in the newly collected data on the other (see Figure 21). As for the overall number of *-i-* stem verbs, there are 39 *-i-* stem verbs with an attested perfective marker *-sard-* and 23 *-i-* stem verbs with an attested perfective marker *-nd-*. Their proportion is not as high, only 1.7:1, but the dominance of /r/ is still obvious. If we look at the occurrences of the perfective markers in the newly collected data, the proportion is even higher (2.3:1), so the dominance is even more conspicuous.

Now let us take the proportion of consonantal verbs with a stem-final /n/ and a stem-final /v/, which is 1.5:1. This exactly corresponds to the newly collected data, where we find that the proportion of the perfective markers *-nd-* and *-d-* for *-a-* stem verbs is the same, 1.5:1.[82]

82 In Vekerdi (1985), we did not encounter any related data.

6.5.2.2.2 The past forms of verbs with the markers *-av-* and *-ajv-*

As it can be seen in Table 59, the interesting thing about the *-a-* stem verbs is that **they have a variant which corresponds to one of the perfective markers of the consonantal class**: the marker *-j(l)-* can be an equivalent of the marker *-l-*. It also corresponds to the past forms of verbs derived by the marker *-ajv-*. Although only the marker is shared with the consonantal verbs, the derived verbs containing the marker *-ajv-* are also similar in that the marker contains the theme vowel of the *-a-* stem verbs. For *-i-* stem verbs, the phonological form of the marker *-sajl-*, emerging as a perfective marker, is strikingly similar to the past tense form of the marker *ajv-*.

	base	present stem of verb	past stem of verb
consonantal verb	n. a.	*mang-* ‘ask’	*mangl-*
derived consonantal verb	*zōr* ‘strength’	*zorajv-* ‘become strong’	*zoraj(l)-*
-a- stem verb	n. a.	*asá-* ‘laugh’	*asaj(l)-*
derived consonantal verb	*truš* ‘thirst’	*trušajv-* ‘become thirsty’	*trušajl-*
-i- stem verb	n. a.	*bují-* ‘hide’	*bujisajl-*

Table 66

The vocalic verbs and the derived verbs influencing them

The other variant of the past tense of *-a-* stem verbs, with the marker *-d-*, corresponds again to both the consonantal class and a set of verbs containing the derivational marker *-av-*, as mentioned in section 6.5.2.1. The latter one might carry more weight as, just like in the case of *-ajv-*, it actually contains the theme vowel of *-a-* stem verbs.

	base	present stem of verb	past stem of verb
consonantal verb	n. a.	*khel-* ‘dance’	*kheld-*
derived consonantal verb	*ker-* ‘make, do’	*kerav-* ‘have something made/done’	*kerad-*
-a- stem verb	n. a.	*asá-* ‘laugh’	*asad-*

Table 67

The past form of *-a-* stem verbs with the marker /d/ and the possible consonantal influences

The information in Tables 66 and 67 can also be interpreted correlations among schemata. Figure 28 corresponds to Table 66, while Figure 29 corresponds to Table 67.

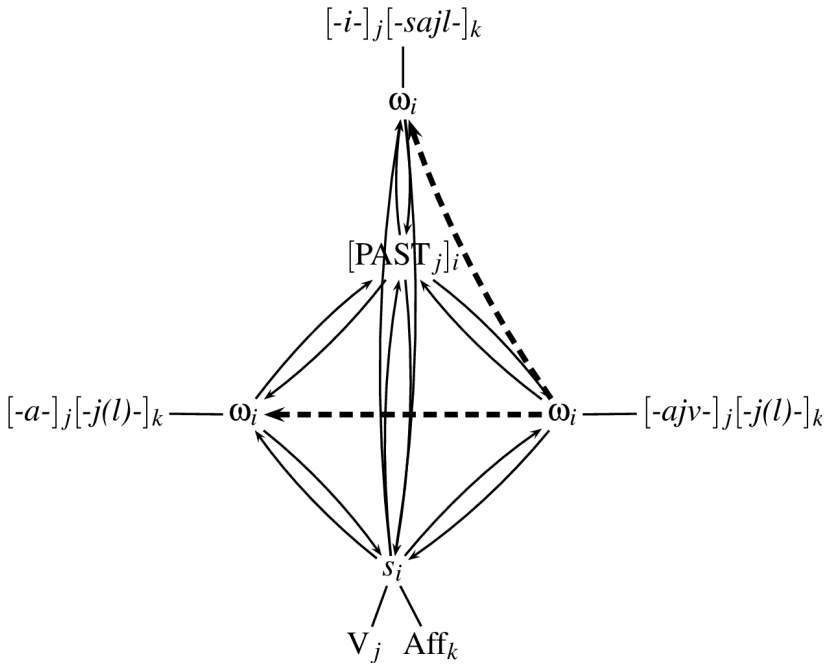

Figure 28

The influence of the derivational marker *-ajv-* on the past tense forms of vocalic verbs

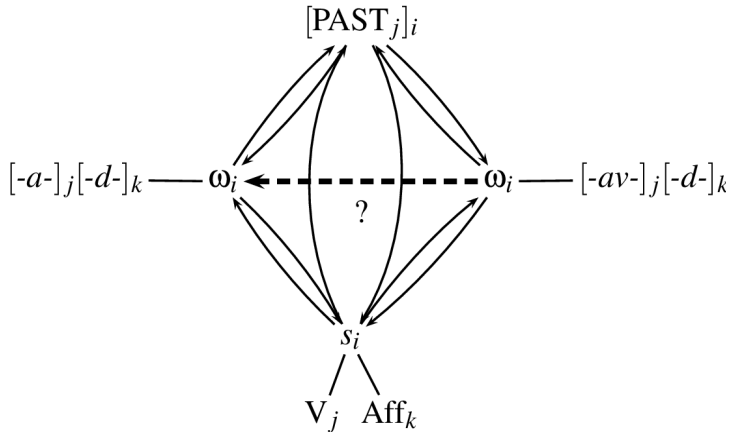

Figure 29

The possible influence of the derivational marker *-av-* on the past tense forms of *-a-* stem verbs

In the newly collected data and Vekerdi (1985), we found 20 attested verbs with the derivational marker *-av-* and 6 with the derivational marker *-ajv-* altogether.

If we compare the proportion of consonantal verbs with a stem-final /r/ and verbs with the derivational marker *-ajv-* to the proportion of the perfective markers *-sard-* and *-sajl-* for *-i-* stem verbs, we find they are fairly close to each other: 5.3:1 versus 6.75:1. We find a similarly high degree of difference if we look at the proportion of the overall number of *-i-* stem verbs with an attested perfective marker *-sard-* and an attested perfective marker *-sajl-* in the newly collected data and in Vekerdi (1985): 3.9:1.

If we look at the *-a-* stem verbs, the results are not so convincing. Based on the type frequency of the verbs with the derivational markers *-av-* and *-ajv-* (their proportion is 3.3:1), the perfective marker *-d-* for *-a-* stem verbs should dominate; however, this is not the case: as we could see in Figure 20, the perfective marker *-j(l)-* dominates in the newly collected data (their proportion is 0.28:1), and the proportions are almost exactly the opposites of each other.

6.5.2.3 Conclusion

In this section, we examined the third weak point in Lovari morphology, the past tense of verbs with either a stem-final /a/ or /i/. After the description of the phenomenon, we looked at two possible aspects that might influence their past forms and found the following.

1. The stem-final /r/, /n/ and /v/ of consonantal verbs. We found that **the proportions of the verbs with these stem-final consonants correspond to the sounds and sound sequences appearing in the past forms of vocalic verbs**, which provides us with evidence that the lack of a straightforward pattern for the vocalic verbs triggers a search for an appropriate pattern, and the most frequent patterns are combinations of inserted sounds and sound sequences which resemble certain, existing markers and contain consonants which clearly reflect the patterns found among consonantal verbs.
2. The past forms of verbs with the derivational markers *-av-* and *-ajv-*. We also looked at the possible connection between the past forms of verbs with the derivational markers *-av-* and *-ajv-* and the past forms of vocalic verbs. **Our results are convincing for the *-i-* stem verbs**, where we compared the proportion of consonantal verbs with a stem-final /r/ and

verbs with the derivational marker *-ajv-* with the proportion of the perfective markers *-sard-* and *-sajl-* for *-i-* stem verbs and found that they were very similar. However, **the results are not so convincing for the *-a-* stem verbs**, where proportion of consonantal verbs with the derivational markers *-av-* and *-ajv-* is quite different from the proportion of the perfective markers *-d-* and *-j(l)-* for the *-a-* stem verbs. This means that this question will definitely need further investigation.

7 Conclusion

7.1 General remarks

Through the example of the Lovari dialect of the Romani language I attempted to demonstrate that variation is an essential part of language and that its study brings us closer to a better understanding of the nature of language change, language acquisition and the essential cognitive processes behind the structure and use of language. Variation, which is present across all dimensions of language, as well as the gradient nature of a high number of linguistic phenomena show the two complementary tendencies in language and, more generally, in human cognition: regularisation on the one hand, which is the attempt to reach a state where mental processing requires the least possible amount of energy; and differentiation on the other, whose aim is to achieve a state where individual items are maximally distinguishable from each other. The simultaneous presence and effects of these two forces make sure that language as an instance of human cognition and individual languages are in constant motion: they keep changing, but in fairly regular ways; there are different patterns within a language, but the overall structure of the patterns is fairly similar; the surface forms differ to a great extent across languages but their functions bear fairly identifiable common traits.

For the study of variation and gradience, analogy proves a good tool. Stochastic rules might work when we would like to predict something that does not exist (a classic experimental example is wug-testing); but, as shown above through Romani, languages often exhibit phenomena where the processes take place in the domain of existing items (forms, paradigms etc.), where the items show uniformity on one level but show variation

on another. We have seen extreme examples of this in the verbal morphology of Lovari, where the same lexical item has only got one realisation in one paradigm but several realisations in another paradigm.

The phenomena we encounter in the nominal and verbal morphology of Lovari show that even variation can be gradient. Within the nominal morphology, we see two distinct, internally uniform patterns for both the masculine oblique forms and the feminine oblique plural forms. On the one hand, uniformity means that we do not find mixed paradigms (nothing in the newly collected data suggests that it is possible for a masculine noun to have *-es-* in the singular oblique and *-on-* in the plural oblique). **On the other, uniformity also refers to what we called regularisation above**: the presence of the marker *-en-* in the feminine plural oblique is variation in the feminine plural paradigms but uniformity in the wider category of nouns. **This is what we might call gradience in the variation in a broader sense.** A matrix of variation in the nominal morphology can be seen in Figure 30. The rectangles contain the two nominal classes, the ellipses contain the possible oblique endings, and the circle represents the only masculine nominal ending affected by variation. The colour blue is neutral, but only two blue elements can be part of any relationship on the matrix at the same time. The red element is not passable, while the green elements are.

The two nominal classes are interrelated through the plural oblique ending *-en-*, while the two possible sets of masculine oblique endings are connected through the nominal ending *-o*. The matrix, through the arrangement of the lines, also shows that all the elements are related somehow, and the more distant they are as we go along the lines, the less closely related they are. Thus, two neighbouring elements, like the masculine class and the ending *-es-* are directly related, while the feminine class and the ending *-o* are only more loosely related: there is no intermediate connection (there is no lexical item where they would appear simultaneously), they only connect through the ending *-en-*. This means that there are lexical items which belong both to the group represented by the ending *-en-* and to the feminine class on the one hand, and there are lexical items which belong both to the group of nouns ending in *-o* and the group represented by by the ending *-en-* on the other, but there are no lexical items belonging to the group of nouns ending in *-o* and the group represented by the ending *-en-* at the same time.

The relationship can be of a different nature, as well; for example, the endings

***-es-* and *-os-* are related through the ending *-o*,** but this relationship is not characterised by simultaneity, rather by a kind of correspondence or interchangeability.

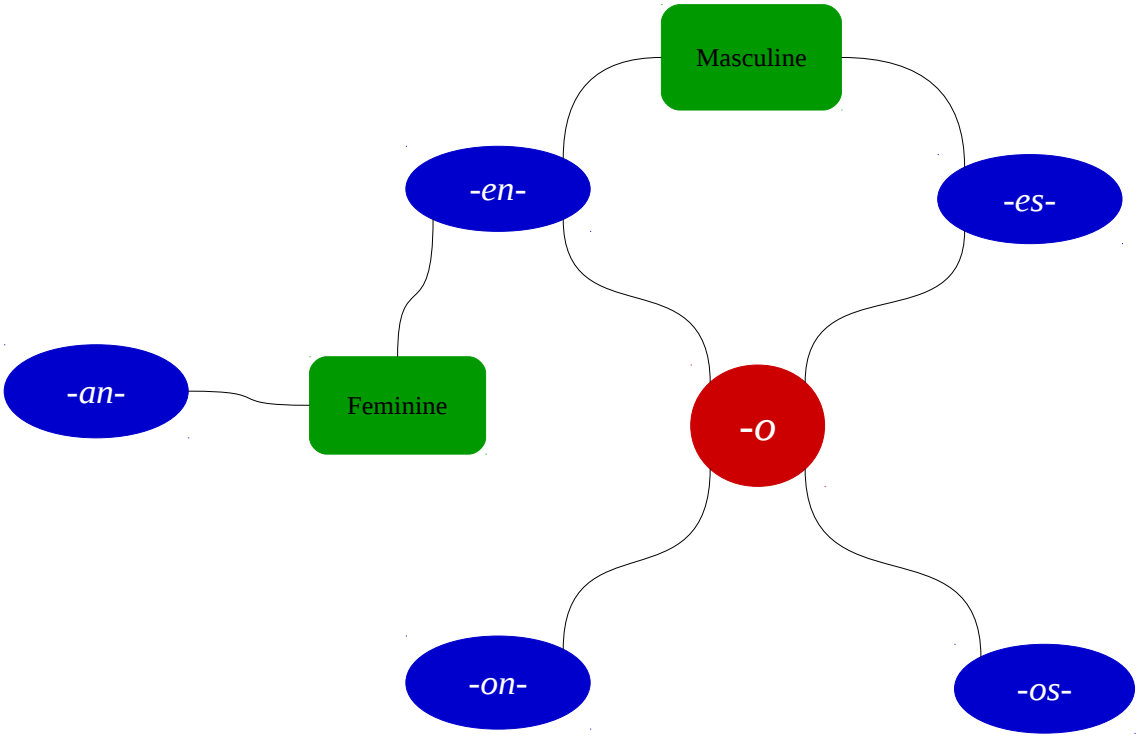

Figure 30

Variation in the nominal morphology of Lovari

In a narrower sense, gradience in the variation can refer to intra-categorial phenomena, like in the case of the past tense of verbs. For consonantal verbs, the perfective marker depends entirely on the stem-final consonant. Uniformity like that is not present among the *-a-* stem verbs, where the stem-final vowel cannot predict the perfective marker unambiguously. The marker *-nd-* introduces more uncertainty into the system, especially if we consider that the element /n/ resembles a derivational marker. For the *-i-* stem verbs, regularity and uniformity is further weakened, and variation is even more robust, with all the “regular” perfective markers eliminated from the past tense.

Whereas the past tense is one dimension, we can also see the gradience of variation along another dimension, that of the verb classes. Both the present and past tense of consonantal verbs is fairly regular. If we move onto the *-a-* stem verbs, we can see that while the same regularity applies to the present tense, variation emerges in the past tense. For *-i-* stem verbs, variation already occurs in the present tense, see for example the first person plural forms. A matrix of variation in the verbal morphology can be seen in Figure 31.

Here, the blue ellipses represent the possible past tense markers (and a related derivational marker, *-sajl-*), while the red circles contain the verb classes. The single lines connect the verb classes to the possible markers, while the dotted lines represent interrelations between any two of the markers themselves. The two possible markers for the consonantal class are not interchangeable, hence the colour red (it is not passable), but in the case of the vocalic classes the markers vary, hence the colour green.

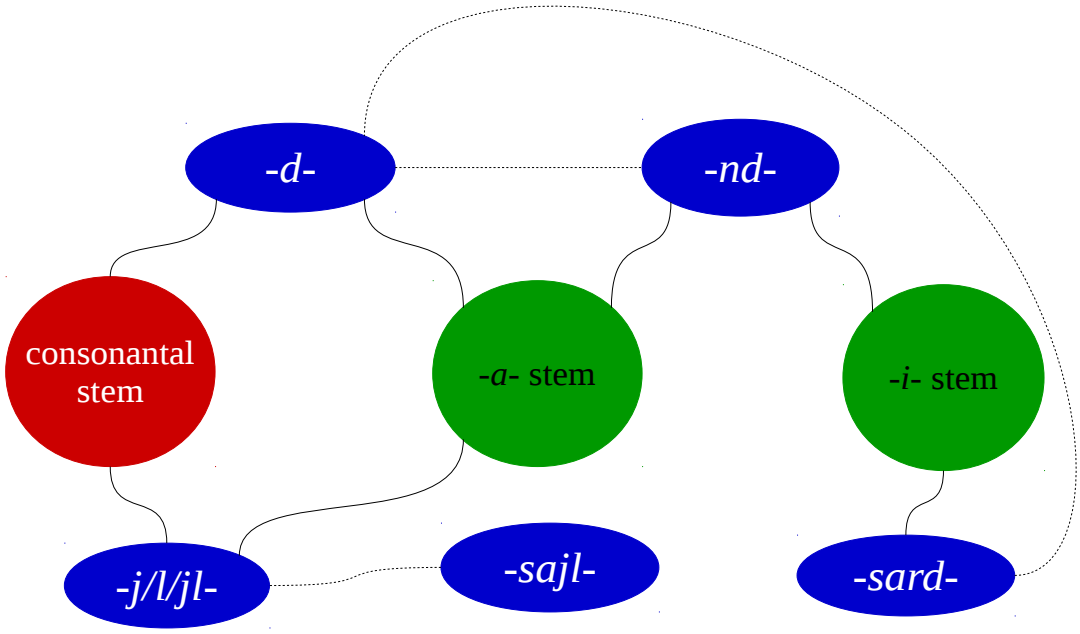

Figure 31

Variation in the verbal morphology of Lovari

7.2 Ongoing research and further points of investigation

A small-size corpus of the Romani varieties spoken in Hungary, including but not limited to Lovari is under preparation within the framework of the project Variation in Romani Morphology, supported by the Hungarian Scientific Research Fund (OTKA, Project 111961, project leader: László Kálmán). **A theoretical question that arises in connection with that is whether the different forms of the same lexical item across the whole of Romani can be called variation or not.** For example, the Lovari (and more generally, Vlax Romani) vocalic verb *loší-* 'be glad' corresponds to the consonantal verb *lošan-* in Romungro. Is it a stem where there are two varying forms, or should we say that the two dialects are far enough from each other to consider them separately?

As the corpus increases, we are expecting to be able to **extract data concerning token frequency** as well, which will hopefully confirm our assumptions. We would also like to examine the data with algorithms and analogical models as soon as it becomes possible.

The contrast between the proportions of the verbs with the derivational markers *-av-* and *-ajv-* and *-a-* stem verbs with the perfective markers *-d-* and *-j(l)-* certainly needs to be elaborated further; it is possible that, in spite of the higher type frequency of causative verbs, the intransitive verbs with the derivational marker *-ajv-* have higher token frequency, which would tip the scales in favour of the perfective marker *-j(l)-*.

Romani also offers a plethora of other phenomena to look into. Among others, we would like to further elaborate on the following aspects:

1. **multiple exponence**, that is, multiple morphological realisations of a single feature (as described, for example, in Caballero & Harris 2012);
2. **syntactic and morphological constructions** in environments of language contact;
3. **morphological recycling** and thus, the rearrangement of surface forms into different patterns.

Appendix

The original recordings and their transcriptions can be found at

http://www.nytud.hu/oszt/elmnyelv/balo/dissz/recordings.zip
and
http://www.nytud.hu/oszt/elmnyelv/balo/dissz/transcriptions.pdf.

The original Hungarian sentences of the questionnaire follow below.

1) Itt élek, és a szüleim is itt éltek.
2) Megdöglik a lovam, ha nem adok neki enni.
3) Imádkozol értem, édesanyám?
4) A férfiak előbújtak a sátrakból.
5) Elcserélem a régi kocsimat egy újra.
6) A fiam most nincs itthon, de mindennap írok neki levelet.
7) Amikor felnövünk, mi is dolgozni fogunk.
8) Azokkal a késekkel nem tudsz kenyeret vágni.
9) Az a lány tetszik nekem.
10) Te milyenre cseréled a telefonod?
11) Egész nap a mobiljával szórakozik.
12) Átkokat zúdított a fejemre.
13) Ezekben az utcákban nem laknak cigányok.
14) A fiú a fűben fekszik.
15) Cigány kenyeret készítek vacsorára.
16) Hozd ide az abroncsokat!
17) Borotvával vágta el a szomszédja torkát.
18) Meglátta a koldust és elszaladt.
19) Elegem van a legyekből.
20) Mindenre rászállnak ezek a rohadt legyek.
21) Menj ki innen, mert megharagszom.
22) Késsel szúrta le a szomszédját.

23) Addig gúnyolta, amíg az megütötte.

24) Az ágyakat és a székeket kivittük az udvarra.

25) A vonatok mindig késnek.

26) Nem bírom megenni, olyan sok.

27) Nem mert kimenni az utcára.

28) Mi van azokban a poharakban?

29) Vizet húzott fel a kútból egy vödörben.

30) Az asztalokon sok étel és bor volt.

31) Hány óra van?

32) Nem szeretjük az idegeneket.

33) Megnősz és nagy leszel!

34) A pap megáldotta a fiatalokat.

35) A füvek szaga száll a levegőben.

36) Találkoztok boldog cigányokkal?

37) Tűket tett a fenekem alá.

38) Miért bántasz?

39) A király megvakítja a rossz embereket.

40) Nem hittünk az embernek, aki pénzt ígért nekünk.

41) A lány ránevetett a fiúra.

42) Adj enni a libáknak.

43) Valaki rálépett az órámra, és összetört.

44) Milyen dezodort használsz?

45) Nem mertük azt mondani, hogy ne jöjjenek.

46) Az abroncsban van egy lyuk.

47) Amikor jöttem ide, koldusokkal találkoztam.

48) Most kellesz neki, vagy nem?

49) Mikor kezdődik a tévében a műsor?

50) Tagadod, hogy éjszaka bemásztál a kertembe?

51) Mindennap megetetjük az állatokat.

52) Ha tanulunk, előrébb jutunk.

53) Annak a fának az ágaival mindig csak a baj van.

54) A családoknak több pénz kellene.

55) Elbújok, és te megkeresel, jó?

56) Levelet írtam a királynak.

57) A kocsmárosokkal mindig van valami baj.

58) Ne a ruhád ujjaival töröld le az asztalt!

59) Segítségért kiáltottak.

60) Ki fog minket kihúzni a bajból?

61) A jégesők elverték idén a búzát.

62) Mindig kiszabadítjuk, amikor elkapják.

63) Lefekszünk a földre és alszunk egyet.

64) Mindig vannak gondok, de mindig rendbe jönnek.

65) Ne nézz oda, figyelnek minket.

66) Nagyon sokba kerülnek a laptopok.

67) Sokáig ápoltuk a beteg nagymamánkat.

68) Mondták az apósoknak, hogy jó volt az esküvő.

69) Alig tudtunk menni a hóban, fagyban.

70) Vidéken sok embernek nincs munkája és éhezik.

71) A városok nagyon piszkosak.

72) Van valami furcsa abban az állatban.

73) Kértem az idegentől egy cigarettát.

74) Ollóval vágtam el, nem késsel.

75) Adott nekem is a dohányból.

76) Csapd már le azt a hangyát!

77) Emlékeztek, amikor elromlott a számítógép?

78) Kiszabadították a börtönből.

79) Ne csak a gyűrűktől várd a jó házasságot.

80) A kerteknek sok víz kell.

81) Olyan betegek lettünk, egész nap köhögtünk.

82) Kirázom a takarót, mert piszkos.

83) Az emberek kint álltak a kapuikban.

84) Ne vegyél el semmit a kurváktól.

85) Belerohadok ebbe a munkába.

86) Okos király vagyok, okosan uralkodom.

87) Azokkal az ollókkal semmit nem lehet elvágni, olyan életlenek.

88) Senki sincs az utcákon.

89) A boszorkány megátkozza az idegent.

90) A tündér egy felhőn ült.

91) A felhők az égen nagyon szépek.

92) Utálom az édességet.

93) Szépek a mezők virágai.

94) Vigyázz, ömlik a víz!

95) Amikor veled vagyok, megfiatalodom.

96) Régen sátrakban laktak az emberek.

97) Megvakulok, olyan nagyon süt a nap.

98) Te milyenre cserélted a telefonod?

99) Tele lett a kert ágakkal és levelekkel.

100) Elmentem a családdal a városba.

101) A pap megáldja a fiatalokat.

102) Sajnálom, de nem tudok segíteni.

103) Amikor idejön, mindig csak mobilozik.

104) Tele van szúnyogokkal a szoba.

105) Van valami kosz a ruhád ujjain.

106) Szél fúj a mezőkön.

107) Most beteg az apósom, kórházban ápolják.

108) A kutak vize itt nem jó.

109) A vonat mindjárt indul.

110) Amikor megöregszünk, átadjuk a helyünket a fiatalságnak.

111) Bántottuk őt, de ő is bántott minket.

112) Semmi bajom a hangyákkal.

113) A köldökében ékszert hord.

114) Megszúrtam a tűvel az ujjam.

115) Hála istennek javul az idő.

116) Amikor megérkezem, kiáltok.

117) A pokrócokat ki kell mosni, mert piszkosak.

118) Olyan beteg lett, egész nap köhögött.

119) A fejéhez kapott.

120) Elő a poharakkal!

121) Elbújok az asztal alá.

122) Elszakadnak az ingeim, mert nem vigyázok rájuk.

123) Megbolondulok már, annyi hülyeséget beszélsz.

124) Megszökött a börtönből.

125) Szépek a gyűrűk az ujjaidon.

126) Becsomagoljuk a ruhánkat és elmegyünk máshova.

127) Én is tetszettem a lánynak.

128) Kérd el a menyasszonytól a gyűrűt.

129) Nem bírt elfutni.

130) Van azoknak a falatoknak íze?

131) Nem bánok semmit.

132) A mennykő essen beléd!

133) Egész életében sántított.

134) Sokat beszélgetünk az életről.

135) A pálinka megerősíti az embert, testben és lélekben.

136) A kurvák élete nehéz.

137) A kocsmárostól kértem egy sört, mert szomjas voltam.

138) Pokróccal takarózom, mert nincs paplanom.

139) Egy hosszú nap után mindig elfáradok.

140) Mire gondolsz, barátom?

141) Szögre akasztom a hegedűm.

142) Az idők végén minden bűnünket megbocsátja Isten.

143) Mit nyertek, ha ti győztök a versenyen?

144) A pálinkától az ember megerősödik, testben és lélekben.

145) Elkezdjük a játékot, és majd jöttök, amikor tudtok.

146) Mindig megszökik a börtönből, amikor elkapják.

147) Odaadtam az apósnak a pénzt.

148) Kalapácsokat hoztam tegnap a szomszédból.

149) Sokat gondolkoztunk, hogy mit tegyünk.

150) Úgy megrúgta, hogy nem tudott felállni.

151) Okos király volt, okosan uralkodott.

152) Vonattal jöttem, nem busszal.

153) Rosszul lettem a hírtől.

154) Finom a dohány szaga.

155) Megörültem, amikor hallottam, hogy esküvő lesz.

156) Eljöttem az asztaloktól.

157) A fazekak fülei melegek.

158) Még ebből a kicsi falatból is el akarsz venni?

159) Valami mindig elromlik.

160) A tanító mindig segít az embereknek.

161) A bajokban mindig van valami jó is.

162) Ha a tévébe az kell, akkor lefogyunk.

163) A ház falai még állnak, de az ajtókat már kiszedték.

164) Ásd el a földbe.

165) Nem látszunk ezen a képen.

166) Miért nem bízol az idegenekben?

167) Örültem az esküvőnek.

168) A fiatalokkal megyünk moziba.

169) Sokat beszélgettünk az életről.

170) Felakasztottam a kabátom a szögre.

171) Mindenhova a laptopommal megyek.

172) Megfürdök és utána elmegyek lefeküdni.

173) A falu kertjeiben sok szép fa nő.

174) Nem gondoltam, hogy lopott.

175) Imádkoztál értem, édesanyám?

176) A menyasszonyokat sokáig kell öltöztetni.

177) Amikor a fiam nem volt itthon, mindennap írtam neki levelet.

178) Miközben ástunk, a földből előkerült egy láda.

179) Min gondolkozol, barátom?

180) Az orvosok életeket mentenek.

181) A falba verem a fejem.

182) Hittek nekünk, pedig nem mondtunk igazat.

183) Mindig csak rossz dolgokat mondanak a hírekben.

184) A találkozón a királyoknak adták a legjobb helyeket.

185) Fő az étel a fazékban.

186) A borotvákkal vigyázni kell, mert élesek.

187) Amikor idejött, mindig csak mobilozott.

188) A kapukból kiestek a csavarok.

189) Találkozom boldog cigányokkal.

190) Megjavítod a kocsit a jövő héten?

191) Bocsássatok meg nekünk, barátaim.

192) A boszorkány megátkozta az idegent.

193) Bánatomban sírok.

194) A macska a farkával játszott.

195) Gondtól gondig tart az életünk.

196) A szúnyogok csak szívják a vérünket.

197) Az átkokkal nem érsz el semmit.

198) Leesett a kalapács feje.

199) A fiataloknak adjuk a kocsinkat.

200) Mobilokat adok-veszek, ezzel keresem a pénzt.

201) Segítettünk a szegényeknek.

202) Elcseréltem a régi kocsimat egy újra.

203) Csak az ágyban fekszik egész nap.

204) Egy faluban lakom, de a városban születtem.

References

Achim, Viorel (2004). *The Roma in Romanian History.* Budapest: Central European University Press.

Adamou, Evangelia (2016). *A Corpus-Driven Approach to Language Contact.* Berlin: Mouton de Gruyter.

Ackerley, Frederick G. (1914). The Romani speech of Catalonia. *Journal of the Gypsy Lore Society, new series* **8**. 99–140.

— (1929). Basque Romani. *Journal of the Gypsy Lore Society, third series* **8**. 50–94.

— (1932). A Lovari vocabulary. *Journal of the Gypsy Lore Society, third series* **11**. 124–187.

Ackerman, Farrell & Robert Malouf (2016). Implicative Relations in Word-Based Morphological Systems. In Andrew Hippisley & Gregory Stump (eds.) *The Cambridge Handbook of Morphology*. Cambridge: Cambridge University Press. 297–327.

Ackerman, Farrell, James P. Blevins & Robert Malouf (2009). Parts and Wholes: Implicative Patterns in Inflectional Paradigms. In Blevins & Blevins (2009). 54–82.

Aha, David W., Dennis Kibler & Marc K. Albert (1991). Instance-based learning algorithms. *Machine Learning* **6**. 37–66.

Albright, Adam (2009). Modeling analogy as probabilistic grammar. In Blevins & Blevins (eds.) (2009). 185–213.

Albright, Adam & Bruce Hayes (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* **90:2**. 119–161.

Allen, Cynthia L. (1995). *Case Marking and Reanalyis*. Oxford: Oxford University Press.

Anttila, Raimo (2002). Variation and Phonological Theory. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Oxford: Blackwell. 206–243.

Arató, Mátyás (2015). *A dél-baranyai muncsán dialektus nyelvi rendszere*. MA dissertation. Pécs: Pécsi Tudományegyetem.

Auwera, Johan van der (1998). Phasal adverbials in the languages of Europe. In Johan van der Auwera (ed.) *Adverbial Constructions in the Languages of Europe.* Berlin: Mouton de Gruyter. 25 –145.

Baayen. Harald (1992). On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.) *Yearbook of Morphology*. Dordrecht: Kluwer Academic. 181–208.

— (2003). Probabilistic approaches to morphology. In Bod *et al.* (eds.) (2003). 229–287.

Bakker, Peter (1995). Notes on the genesis of Caló and other Iberian Para-Romani varieties. In Matras (ed.) (1995). 125–150.

— (1997). Athematic morphology in Romani: The borrowing of a borrowing pattern. In Matras *et al.* (eds.) (1997). 1–21.

— (1998). Para-Romani language versus secret languages: differences in origin, structure and use. In Matras (ed.) (1998). 69–96.

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

Bakker, Peter & Yaron Matras (1997). Introduction. In Matras *et al.* (eds.) (1997). VII–XXX.

Bakker, Peter & Marcia Rooker (2001). *The political status of the Romani language in Europe*. Mercator. Dret i Legislació Lingüistics. Ciemen Working Papers 3.

Bakker, Peter & Hein van der Voort (1991). Para-Romani languages: an overview and some speculations on their genesis. In Peter Bakker & Marcel Courthiade (eds.) (1991). 16–44.

Bakker, Peter, Aymeric Daval Markussen, Mikael Parkvall & Ingo Plag (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* **26:1**. 5–42.

Bakker, Peter & Marcel Courthiade (eds.) (1991). *In the Margin of Romani. Gypsy Langauges in Contact.* Amsterdam: Institute for General Linguistics.

Baló, Márton András (2008). A Strange Case of Defectiveness in the Lovari Verbal Paradigm. In László Kálmán (ed.) *Papers from the Mókus Conference*. Budapest: Tinta. 118–136.

— (2011). A current trend or a historic remnant? The case of a Lovari verb-forming suffix. P*oznań Studies in Contemporary Linguistics* **47**. 264–282.

— (2012). Arguments from Lovari loan-verb adaptation for an analogy-based analysis of verbal systems. In Kiefer *et al.* (eds.) (2012). 3–22.

— (2015). The nominal morphology of Lovari from an analogical perspective. In *Acta*

Linguistica Hungarica **62:4**. 395–414.

Bari, Károly (1990). *Le vēšeski dēj. Az erdő anyja*. Budapest: Országos Közművelődési Központ.

Bartha, Csilla (ed.) (2007). *Cigány nyelvek és közösségek a Kárpát-medencében.* Budapest: Nemzeti Tankönyvkiadó.

Bauer, Laurie (1988). *Introducing linguistic morphology.* Edinburgh: Edinburgh University Press.

Beníšek, Michael (2012). On the origin of the consonantal feminine stems in Romani. Paper presented at the 10th International Conference on Romani Linguistics, Barcelona.

Blake, Barry J. (2000). Case. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.) *Morphologie/Morphology. Ein internationales Handbuch zur Flexion und Wortbildung/An International Handbook on Inflection and Word-Formation.* Berlin: Mouton De Gruyter. 1073–1090.

— (2001). *Case*. Cambridge: Cambridge University Press.

Blevins, James P. (2006). Word-based morphology. *Journal of Linguistics* **42:3**. 531–573.

— (2013). Word-based morphology from Aristotle to modern WP. In Keith Allan (ed.) *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press. 375–396.

Blevins, James P. & Juliette Blevins (eds.) (2009). *Analogy in Grammar: Form and Acquisition.* Oxford: Oxford University Press.

Bloomfield, Leonard (1933). *Language*. New York: Holt.

Boas, Hans (2003). *A constructional approach to resultatives. Volume 14 of Stanford monographs in linguistics.* Stanford, CA: CSLI Publications.

Bod, Rens Jennifer Hay & Stefanie Jannedy (eds.) (2003). *Probabilistic Linguistics*. Cambridge, MA: The MIT Press.

Bodnárová, Zuzana (2013). Overview of Vend Romani in Hungary. In Barbara Schrammel-Leber & Barbara Tiefenbacher (eds.) *Romani V. Papers from the Annual Meeting of the Gypsy Lore Society.* Graz: Grazer Romani Publikationen. 29–41.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* **32:1**. 45–86.

Bonami, Olivier & Gilles Boyé (2006). Deriving Inflectional Irregularity. In Stefan Müller

(ed.) *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna*. Stanford, CA: CSLI Publications. 361–380.

Booij, Geert (2007). Construction Morphology and the Lexicon. In Fabio Montermini, Gilles Boyé & Nabile Hathout (eds.) *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville: Cascadilla Proceedings Project. 34–44.

— (2008). Recycling Morphology: Case Endings as Markers of Dutch Constructions. Leiden University, Leiden. Available (June 2015) at http://www.academia.edu/16819956/Recycling_morphology_Case_endings_as_markers_of_Dutch_constructions.

— (2010). *Construction Morphology.* Oxford: Oxford University Press.

Boorde, Andrew (1870). *The fyrst boke of the introduction of knowledge made by Andrew Borde, of physycke doctor. A compendyous regyment, or, A dyetary of helth made in Mountpyllier.* London: Published for the Early English Text Society by N.T. Trübner & Co.

Boretzky, Norbert (1985). Sind Zigeunersprachen Kreols? In Norbert Boretzky, Werner Enninger & Thomas Stolz (eds.) *Akten des 1. Essener Kolloquiums über Kreolsprachen und Sprachkontakt*. Bochum: Brockmeyer. 43–70.

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

— (1992). Zum Erbwortschatz des Romani. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **45**. 227–251.

— (1994). *Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar.* Wiesbaden: Harrassowitz.

— (1995a). Armensiches im Zigeunerischen (Romani und Lomavren). *Indogermanische Forschungen* **100**. 137–155.

— (1995b). Interdialectal interference in Romani. In Matras (ed.) (1995). 69–94.

— (1998). Der Romani-Wortschatz in den Romani-Misch-Dialekten (Pararomani). In Matras (ed.) (1998) 97–132.

— (2003). *Die Vlach-Dialekte des Romani*. Wiesbaden: Harrassowitz.

Boretzky, Norbert & Birgit Igla (1991). *Morphologische Entlehnung in den Romani-Dialekten. (Arbeitspapiere des Projektes 'Prinzipien des Sprachwandels' 4).* Essen:

Universität GH Essen.

Bubeník, Vít (2000). Was Proto-Romani an ergative language? In Elšík & Matras (eds.) (2000) 205–227.

Bybee, Joan (1985). *Morphology: a study of the relation between meaning and form.* Amsterdam: John Benjamins.

— (1995). Regular Morphology and the Lexicon. *Language and Cognitive Processes* **10:5**. 425–455.

— (1998). The Emergent Lexicon. In M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson & Tamra Wysocki (eds.) *CLS 34: The panels.* Chicago: Chicago Linguistics Society. 421–435.

— (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

— (2006). From usage to grammar: the mind's response to repetition. *Language* **82**. 711–733.

— (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & David Eddington (2006). A usage-based approach to Spanish verbs of 'becoming'. *Language* **82**. 323–355.

Bybee, Joan & Paul Hopper (2001). *Frequency effects and the emergence of linguistic structure.* Amsterdam: John Benjamins.

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

Bybee, Joan & William Pagliuca (1987). The evolution of future meaning. In Anna Giacalone Ramat, Onofrio Carruba & Giuliano Bernini (eds.) *Papers from the 7th International Conference on Historical Linguistics.* Amsterdam: John Benjamins. 109–122.

Bybee, Joan, Revere Perkins & William Pagliuca (1994). *The evolution of grammar: tense, aspect and modality in the language of the world.* Chicago: University of Chicago Press.

Bybee, Joan & Joanne Scheibman (1999). The Effect of Usage on Degrees of Constituency: The Reduction of Don't in English. *Linguistics: An Interdisciplinary Journal of the Language Sciences* **37:4**. 575–596.

Caballero, Gabriela & Alice C. Harris (2012). A working typology of multiple exponence. In Kiefer *et al.* (eds.) (2012). 163–188.

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

Chandler, Steve (1998). Instance-based reference for past-tense verb-forms: An experimental study. Paper presented at the First International Conference on the Mental Lexicon, Edmonton, Alberta, Canada.

— (2002). Skousen's analogical approach as an exemplar-based model of categorization. In Royal Skousen, Deryle Lonsdale & Dilworth B. Parkinson (eds.) *Analogical Modeling. An exemplar-based approach to language.* Amsterdam: John Benjamins. 51–105.

Choli-Daróczi, József & Levente Feyér (1988). *Cigány nyelvkönyv.* Budapest: Magyarországi Cigányok Kulturális Szövetsége.

Chomsky, Noam (1965). *Aspects of the theory of syntax.* Cambridge, MA: The MIT Press.

— (1980). *Rules and representations.* Oxford: Blackwell.

Cohn, Abigal C., Cécile Fougeron & Marie K. Huffman (eds.) (2012). *The Oxford Handbook of Laboratory Phonology.* Oxford: Oxford University Press.

Cost, Scott & Steven Salzberg (1993). A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features. *Machine Learning* **10**. 57–78.

Costanzo, Angelo (2008). On the expansion and productivity of a Balkan Romance conjugational sub-class. Paper presented at the First Oxford Workshop on Romance Verb Morphology, Oxford.

Courthiade, Marcel (1989). La langue Romani (Tsigane): évolution, standardisation, unification, réforme. In István Fodor & Claude Hagège (eds.) *Language Reform. History and Future. Volume IV.* Hamburg: Helmut Buske. 87–110.

— (1991). Romani versus Para-Romani. In Bakker & Courthiade (eds.) (1991). 1–15.

Cser, András. 2014. Fonologizálódás, morfologizálódás, lexikalizálódás: fejezetek a hangváltozások életéből. *Nyelvtudományi Közlemények* **110**. 319–332.

Daelemans, Walter & Antal van den Bosch (2005). *Memory-Based Language Processing*.

Cambridge: Cambridge University Press.

Davies, Norman (1996). *Europe: A History.* Oxford: Oxford University Press.

Davies, Anna M. (2004). Saussure and Indo-European Linguistics. In Carol Sanders (ed.) *The Cambridge Companion to Saussure.* Cambridge: Cambridge University Press. 9–29.

Dupcsik, Csaba (2009). *A magyarországi cigányság története.* Budapest: Osiris.

É. Kiss, Katalin. 2010. Valószínűleg, hogy román kontaktushatás. In Katalin É. Kiss & Attila Hegedűs (eds.) *Nyelvelmélet és kontaktológia.* Piliscsaba: PPKE BTK. 223–237.

Eddington, David (1996). The psychological status of phonological analyses. *Linguistica* **36**. 17–37.

— (2000). Analogy and the dual-route model of morphology. *Lingua* **110**. 281–298.

— (2003). Issues in modeling language processing analogically. *Lingua* **114**. 849–871.

— (2006). Paradigm Uniformity and Analogy: The Capitalistic versus Militaristic Debate. *International Journal of English Studies* **6**. 1–18.

Eliav-Feldon, Miriam (2012). *Renaissance Impostors and Proofs of Identity.* London: Palgrave Macmillan.

Elšík, Viktor (1997). Towards a morphology-based typology of Romani. In Matras *et al.* (eds.) (1997). 23–59.

— (2000). Romani nominal paradigms: their structure, diversity, and development. In Elšík & Matras (eds.) (2000). 9–30.

Elšík, Viktor & Yaron Matras (2006). *Markedness and Language Change. The Romani Sample.* The Hague: Mouton de Gruyter.

Elšík, Viktor & Yaron Matras (eds.) (2000). *Grammatical relations in Romani: The noun phrase.* Amsterdam: John Benjamins.

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

Fehértói, Katalin (1987). A középkori cigány helynevek és személynevek. *Névtani Értesítő* **12**. 5–17.

Fennesz-Juhasz, Christiane & Moses F. Heinschink (1999). *Texte österreischer Lovara II. Arbeitsbericht 4 des Projekts Kodifizierung der Romanes-Variante der österreichischen Lovara (hrsgg. von Dieter W. Halwachs).* Vienna: Verein Romano

Centro.

Ferguson, Charles A. (1959). Diglossia. *Word* **15**. 325–340.

Fillmore, Charles J. & Paul Kay (1987). *The goals of Construction Grammar. Berkeley Cognitive Science Program Working Paper 50.* Berkeley, CA: University of California at Berkeley.

Finck, Franz N. (1907). *Die Sprache der armenischen Zigeuner.* St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Fishman, Joshua A. (1967). Bilingualism with and without Diglossia. Diglossia with and without Bilingualism. *The Journal of Social Issues* **2**. 29–38.

Fraser, Angus (1992). *The Gypsies.* Oxford: Blackwell.

Friedman, Victor A. (1991a). Case in Romani: old grammar and new affixes. *Journal of the Gypsy Lore Society, fifth series* **1**. 85–102.

— (1991b). Romani nominal inflection: cases or postpositions? In Maciej Grochowski (ed.) *Problemy opisu gramatycznego języków słowiańskich.* Warsaw: Polska Akademia Nauk, Instytut Języka Polskiego. 57–63.

Garrett, Andrew & Juliette Blevins (2009). Analogical morphophonology. In Kristin Hanson & Sharon Inkelas (eds.) *The Nature of the Word.* Cambridge, MA: The MIT Press. 527–546.

Gilliat-Smith, Bernard J. (1960). Basic Romani. *Journal of the Gypsy Lore Society, third series* **39:1**. 30–34.

Gilsenbach, Reimar (1994). *Weltchronik der Zigeuner. Von den Anfängen bis 1599.* Frankfurt-am-Main: Peter Lang.

Givón, Talmy (1995). *Functionalism and Grammar.* Amsterdam: John Benjamins.

Goldberg, Adele E. (1995). *Constructions.* Chicago: The University of Chicago Press.

— (2006). *Constructions at Work.* Oxford: Oxford University Press.

Goldsmith, John (2009). Morphological analogy: Only a beginning. In Blevins & Blevins (eds.) (2009). 137–163.

Grétsy, László & Gábor Kemény (2005). *Nyelvművelő kéziszótár.* Budapest: Tinta.

Grumet, Joanne (1985). On the genitive in Romani. In Grumet (ed.) (1985). 84–90.

Grumet, Joanne (ed.) (1985). *Papers from the fourth and fifth annual meetings of the Gypsy Lore Society, North American Chapter.* New York: Gypsy Lore Society.

Hajdu, André (1960). Three Romani tales. *Journal of the Gypsy Lore Society, third series*

39:3–4. 92–122.

Halford, Graeme S. & Glenda Andrews (2007). Domain general processes in higher cognition: analogical reasoning, schema induction and capacity limitations. In Maxwell J. Roberts (ed.) *Integrating the Mind: Domain General versus Domain Specific Processes in Higher Cognition.* New York: Psychology Press. 213–232.

Halwachs, Dieter W (1993). Polysystem, Repertoire und Identität. *Grazer Linguistische Studien* **39 –40**. 71–90.

— (1998). *Amaro vakeripe Roman hi – Unsere Sprache ist Roman: Texte, Glossar und Grammatik der burgenländischen Romani-Variante.* Klagenfurt: Drava.

— (2012). Linguistic diversity, dominated languages and the Internet: the case of Romani. In Evgeny Kuzmin & Anastasia Parshakova (ed.) *Linguistic and Cultural Diversity in Cyberspace. Proceedings of the 2nd International Conference.* Paris/Moscow: UNESCO. 272–281.

Hancock, Ian. 1970. Is Anglo-Romanes a creole? *Journal of the Gypsy Lore Society, third series* **49**. 41-44.

— (1984). Romani and Angloromani. In Peter Trudgill (ed.) *Language in the British Isles.* Cambridge: Cambridge University Press. 367–383.

— (1992). The social and linguistic development of Scandoromani. In Ernst H. Jahr (ed). *Language contact: theoretical and empirical studies.* Berlin: Mouton de Gruyter. 37–52.

— (1993). The emergence of a union dialect of North American Vlax Romani, and its implications for an international standard. *International Journal of the Sociology of Language* **99**. 91–104.

— (1995). *A handbook of Vlax Romani.* Columbus: Slavica.

Hare, Mary & Jeffrey L. Elman (1995). Learning and morphological change. *Cognition* **56**. 61–98.

Haspelmath, Martin (1998). Does grammaticalization need reanalysis? *Studies in Language* **22:2**. 315–351.

— (2004). Explaining the Ditransitive Person-Role Constraint: A usage-based approach. In *Constructions* **2**. Available (May 2016) at http://journals.linguisticsociety.org/elanguage/constructions/article/view/3073.html

Hay, Jennifer (2001). Lexical frequency in morphology: is everything relative? *Linguistics*

39. 1040–1070.

Heine, Bernd & Tania Kuteva (2003). *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Herin, Bruno (2012). The Domari Language of Aleppo (Syria). *Linguistic Discovery* **10:2**. Available (April 2016) at http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/412

Hockett, Charles F. (1966). *The state of the art.* The Hague: Mouton.

Holm, John (2000). *An introduction to pidgins and creoles.* Cambridge: Cambridge University Press.

Holzinger, Daniel (1993). *Das Romanes: Grammatik und Diskursanalyse der Sprache der Sinte. (Innsbrucker Beiträge zur Kulturwissenschaft 85).* Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Hopper, Paul & Elizabeth Traugott (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Hübschmannová, Milena (1972). What can sociology suggest about the origin of Roms? *Archiv orientální* **40:1**. 51–64.

Hutterer, Miklós & György Mészáros (1967). *A lovári cigány dialektus leíró nyelvtana.* Budapest: Magyar Nyelvtudományi Társaság.

Humboldt, Wilhelm von (1999). *On Language.* On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species. Cambridge: Cambridge University Press.

Ignasi-Xavier, Adiego (2005). The vestiges of Caló today. In Schrammel *et al.* (eds.) (2005). 60–78.

Itkonen, Esa (2005). *Analogy as Structure and Process.* Amsterdam: John Benjamins.

Jackendoff, Ray. 2008. *Construction after construction* and its theoretical challenges. *Language* **84**: 8–28.

Johansson, Roger (1977). *Svensk Rommani. (Acta Academiae Regiae Gustavi Adolphi 55.).* Uppsala: Lundquist.

Jurafsky, Daniel (2003). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In Bod *et al.* (eds.) (2003). 1–50.

Kálmán, László (2007). Holisztikus szemlélet a nyelvészetben. *Szabad Változók* **4**. Available (November 2015) at http://www.szv.hu/cikkek/holisztikus-szemlelet-a-

nyelveszetben.

Kálmán, László, Péter Rebrus & Miklós Törkenczy (2011). Lehet-e az analógiás nyelvelmélet szinkrón? In Szilágyi & Kádár (eds.) (2011). 31–43.

— (2012). Possible and impossible variation in Hungarian. In Kiefer *et al.* (eds.) (2012). 23–49.

Kaufman, Terrence (1979). Review of Weer Rajendra Rishi, Multilingual Romani Dictionary. *International Journal of the Sociology of Language* **19**. 131–144.

— (1984). Explorations in proto-Gypsy phonology and classification. Paper presented at the session on Romani linguistics at the 6th South Asian Languages Analysis Roundtable, Austin, Texas.

Keine (Stefan). 2007. Reanalysing Hindi Split-Ergativity as a morphological phenomenon. In Trommer, Jochen & Andreas Opitz (eds.) *1 2 many: One to many relations in grammar. (Linguistische Arbeits Berichte 85)*. Leipzig: Institut für Linguistik, Universität Leipzig. 73–127.

Kemény, István (2000). The Structure of Hungarian Roma Groups in Light of Linguistic Changes. *Regio* **2:1**. 105–116.

Kemény, István & Béla Janky (2003). A 2003. évi cigány felmérésről – népesedési, nyelvhasználati és nemzetiségi adatok. *Beszélő* **8:10**. 64–76.

Kenrick, Donald (2007). *Historical Dictionary of the Gypsies (Romanies).* Plymouth: Scarecrow Press.

Kiefer, Ferenc & Mária Ladányi (2000). A szóképzés. In Ferenc Kiefer (ed.) *Strukturális magyar nyelvtan 3. Morfológia.* Budapest: Akadémiai. 137–164.

Kiefer, Ferenc, Péter Siptár & Mária Ladányi (eds.) (2012). *Current Issues in Morphological Theory: (Ir)Regularity, Analogy and Frequency*. Amsterdam: John Benjamins.

Kontra, Miklós (2001). Nyelvi jogok, nyelvi tervezés, és az anyanyelvek tannyelvként való használatának problémái. In Tibor Cserti-Csapó (ed.) Cigány nyelvek nemzetközi szemináriuma (Gypsy Studies – Cigány Tanulmányok 6). Pécs: PTE BTK NTI Romológia Tanszék. 35–45.

Koptjevskaja-Tamm, Maria (2000). Romani genitives in cross-linguistic perspective. In Elšík & Matras (eds.) (2000). 123–149.

Kovalcsik, Katalin & Zsuzsa Kubínyi (2000). *A csenyétei daloskert.* Pécs: Gandhi

Közalapítvány, Gimnázium és Kollégium.

Kraska-Szlenk, Iwona (2007). *Analogy. The Relation between Lexicon and Grammar.* München: Lincom Europa.

Krott, Andrea (2009). The role of analogy for compound words. In Blevins & Blevins (eds.) (2009) 118–136.

Krott, Andrea, Harald Baayen & Robert Schreuder (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics: An Interdisciplinary Journal of Language Sciences* **39:1**. 51–93.

Labov, William (1971). The notion of 'system' in creole languages. In Dell Hymes (ed.) *Pidginization and Creolization of Languages.* Cambridge: Cambridge University Press. 447–472.

— (1972). *Sociolinguistic patterns*. Philadelphia: University of Philadelphia.

— 1982. Building on empirical foundations. In Winfred P. Lehmann & Yakov Malkiel (eds.) *Perspectives on Historical Linguistics*. Amsterdam: John Benjamins. 17–92.

Ladefoged, Peter (1998). Romani elements in non-standard Scandinavian varieties. In Matras (ed.) (1998). 133–161.

Lahiri, Aditi (2000). *Analogy, Levelling, Markedness.* Berlin: Mouton de Gruyter.

Lakoff, George (1982). *Categories and Cognitive Models. Cognitive Science Report No. 2.* Berkeley, CA: Institute of Cognitive Studies, University of California at Berkeley.

Lakoff, George (1987). *Women, fire and dangerous things: what categories reveal about the mind.* Chicago: University of Chicago Press.

Landauer, Attila (2004). Utak és problémák a magyarországi cigánykutatásban. In Attila Nagy & Rita Péterfi (eds.) *A feladatra készülni kell: A cigányság kulturális beilleszkedése és a közkönyvtár.* Budapest: Országos Széchényi Könyvtár – Gondolat. 13–46.

Langacker, Ronald (1987). *Foundations of cognitive grammar. Volume I: Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Langacker, Ronald (1991). *Foundations of cognitive grammar. Volume II: Descriptive applications.* Stanford, CA: Stanford University Press.

Legate, Julie A. (2014). Split ergativity based on nominal type. *Lingua* **148**. 183–212.

Leggio, Daniele V. (2013). *Lace avilen ko radio. Romani language and identity on the Internet.* Ph.D. dissertation, University of Manchester.

Leggio, Daniele V. & Yaron Matras (2013). *Social networks as centres of language codification: Romani on YouTube. RomIdent Working Papers. Paper No. 32.* Available: (March 2016) at http://romani.humanities.manchester.ac.uk/virtuallibrary/librarydb//web/files/pdfs/388/Paper32.pdf.

Lee, Ronald (2005). *Learn Romani: Das-duma Rromanes.* Hatfield: University Of Hertfordshire Press.

Linell, Per (2005). *The Written Language Bias in Linguistics.* Abingdon: Routledge.

Kochanowski, Vania de Gila (1995). Romani language standardization. *Journal of the Gypsy Lore Society, fifth series* **5**. 97–107.

Macalister, Robert A. S. (1914). T*he language of the Nawar of Zutt, the nomad smiths of Palestine. (Gypsy Lore Society Monographs 3).* London: Edinburgh University Press.

Marle, Jaap van (1990). Rule-creating creativity: analogy as a synchronic morphological process. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.) *Contemporary Morphology.* Berlin: Mouton De Gruyter. 267–273.

Masica, Colin P. (1991). *The Indo-Aryan Languages.* Cambridge: Cambridge University Press.

Matras, Yaron (1996). Prozedurale Fusion: grammatische Interferenzschichten im Romanes. *Sprachtypologie und Universalienforschung* **49:1**. 60–78.

— (1998). Para-Romani re-visited. In Matras (ed.) (1998). 1–27.

— (1999). Writing Romani: The Pragmatics of Codification in a Stateless Language. *Applied Linguistics* **20**. 481–502.

— (2002). *Romani: A Linguistic Introduction.* Cambridge: Cambridge University Press.

— (2005a). The classification of Romani dialects: A geographical-historical perspective. In Schrammel *et al.* (eds.) (2005). 7–26.

— (2005b). The Status of Romani in Europe. Report submitted to the Council of Europe's Language Policy Division. Available (December 2013) at http://romani.humanities.manchester.ac.uk/downloads/1/statusofromani.pdf.

— (2010). *Romani in Britain. The Afterlife of a Language.* Edinburgh: Edinburgh University Press.

— (2011). Romani. In Bernd Kortmann & Johan van der Auwera (eds.) *The languages and linguistics of Europe. A comprehensive guide.* Berlin: Mouton de Gruyter. 257–268.

— (2012). *A Grammar of Domari.* Berlin: Mouton de Gruyter.

— (2015). Transnational policy and ‘authenticity’ discourses on Romani language and identity. *Language in Society* **44:3**. 295–316.

Matras, Yaron (ed.) (1995). *Romani in contact: the history and sociology of a language*. Amsterdam: John Benjamins.

— (1998). *The Romani element in non-standard speech*. Wiesbaden: Harrassowitz.

Matras, Yaron & Peter Bakker (2003). The study of mixed languages. In Yaron Matras & Peter Bakker (eds.) *The Mixed Language Debate.* Berlin: Mouton de Gruyter. 1–20.

Matras, Yaron, Peter Bakker & Hristo Kyuchukov (eds.) (1997). *The Typology and Dialectology of Romani*. Amsterdam: John Benjamins.

McCarthy, John J. (2008). *Doing Optimality Theory.* Oxford: Blackwell.

McLane, Merrill (1985). Romani speech domains in Spain and Portugal. In Grumet (ed.) (1985). 188–198.

McWhorter. John H. (2005). *Defining Creole.* Oxford: Oxford University Press.

Mel'čuk, Igor A. (1986). Toward a Definition of Case. In Richard D. Brecht & James S. Levine (eds.) *Case in Slavic*. Columbus: Slavica. 35–85.

Menn, Lise & Cecily Jill Duffield (2014). Looking for a “Gold Standard” to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. In Frederick J. Newmeyer & Laurel B. Preston (eds.) *Measuring Grammatical Complexity.* Oxford: Oxford University Press. 281–302.

Mészáros, György (1968). Lovari-Texte aus Ungarn. *Acta Linguistica Academiae Scientiarium Hungaricae* **18:1–2**. 173–190.

— (1969-1970). Cigánydialektusok Magyarországon. Nyelvföldrajz. In László Mándoki, György Fancsovits, Éva Hárs, Attila Kiss & Anna Vass (eds.) *A Janus Pannonius Múzeum évkönyve XIV-XV.* Pécs: Baranyai Megyei Múzeumok Igazgatósága. 309–319.

— (1976). The Cerhari Gypsy dialect. *Acta Linguistica Academiae Scientiarium Hungaricae* **30:3**. 351–367.

— (1980). *A magyarországi szinto cigányok.* Budapest: Magyar Nyelvtudományi Társaság.

Miklosich, Franz (1872-80). *Über die Mundarten und Wanderungen der Zigeuner Europas.* Vienna: Karl Gerold’s Sohn.

Milroy, James (1992). *Linguistic Variation and Change.* Oxford: Blackwell.

Motsch, Wolfgang (1988). On inactivity, productivity and analogy in derivational processes. *Linguistische Studien* **179**. 1–30.

Nagy, Pál (1998). *A magyarországi cigányok története a rendi társadalom korában.* Kaposvár: Csokonai Vitéz Mihály Tanítóképző Főiskola.

— (2004). *A magyarországi cigányok korai története.* Pécs: PTE BTK Romológia és Nevelésszociológia Tanszék.

Orsós, Anna & László Kálmán (2009). *Beás nyelvtan*. Budapest: Tinta.

Paul, Hermann (1891). *Principles of the History of Language.* London: Longmans, Green and Co.

Penn, Derek C., Keith J. Holyoak & Daniel J. Povinelli (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* **31**. 109–78.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee & Hopper (eds.) (2001). 137–157.

— (2002). Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology 7.* Berlin: Mouton de Gruyter. 101–139.

Pisoni, David B. (1997). Some Thoughts on "Normalization" in Speech Perception. In Keith Johnson and John W. Mullennix (eds.) *Talker Variability in Speech Processing.* San Diego: Academic Press. 9–32.

Plank, Frans (1995). (Re-)Introducing Suffixaufnahme. In Frans Plank (ed.) *Double case. Agreement by Suffixaufnahme.* Oxford: Oxford University Press. 3–110.

Phillips, Betty S. (2006). *Word frequency and lexical diffusion.* London: Palgrave, Macmillan.

Pobożniak, Tadeusz (1964). *Grammar of the Lovari dialect. (Prace komisji orientalistycznej 3)*. Kraków: Państwowe wydawnictwo naukowe.

Port, Robert (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology* **25**. 143–170.

— (2010). Rich memory and distributed phonology. *Language Sciences* **32**. 43–55.

Pott, August (1844-45). *Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache*. Halle: Heynemann.

Prince, Alan & Paul Smolensky (2004). *Optimality Theory. Constraint Interaction in Generative Grammar.* Oxford: Blackwell.

Rebrus, Péter & Miklós Törkenczy (2005). Uniformity and contrast in the Hungarian verbal paradigm. In Laura J. Downing, T. Alan Hall & Renate Raffelsiefen (eds.) *Paradigms in Phonological Theory.* Oxford: Oxford University Press. 262–295.

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

Réger, Zita (1974). A lovári-magyar kétnyelvű cigány gyermekek nyelvi problémái az iskoláskor elején. *Nyelvtudományi Közlemények* **76**. 229–255.

— (1990). *Utak a nyelvhez. Nyelvi szocializáció – nyelvi hátrány.* Budapest: Akadémiai.

— (2002). *Cigány gyermekvilág.* Budapest: L'Harmattan.

Rosch, Eleanor H. & Barbara B. Lloyd (1978). *Cognition and categorization.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, Eleanor H. & Catlin B. Mervis (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* **7**. 573–605.

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

Rung, András (2011). *Magyar főnévi alaktani jelenségek analógiás megközelítésben.* Ph.D. dissertation, Eötvös Loránd University.

Salo, Matt T. & Sheila Salo (1986). Gypsy immigration to the United States. In Joanne Grumet (ed.) *Papers from the sixth and seventh annual meetings of the Gypsy Lore Society, North American Chapter.* New York: Gypsy Lore Society. 85–96.

Sampson, John (1927). Notes on Professor R. L. Turner's 'The position of Romani in Indo-Aryan'. *Journal of the Gypsy Lore Society, third series* **6**. 57–68.

Sapir, Edward (1921). *Language.* New York: Harcourt, Brace & World.

Saussure, Ferdinand de (1959). *Course in General Linguistics.* New York: McGraw-Hill Book Company.

Schmid, Anna (2007). The Dom of Hunza (Northern Areas of Pakistan). In Barbara Brower & Barbara R. Johnston (eds.) *Disappearing Peoples?* Walnut Creek, CA: Left Coast Press. 107–127.

Schneider, Klaus P. (2003). *Diminutives in English.* Tübingen: Max Niemeyer.

Schrammel, Barbara, Dieter W. Halwachs & Gerd Ambrosch (eds.) (2005). *General and applied Romani linguistics: Proceedings from the 6th International Conference on Romani Linguistics.* Munich: Lincom Europa.

Silverstein, Michael. (1976). Hierarchy of Features and Ergativity. In Robert M. W. Dixon (ed.) *Grammatical Categories in Australian Languages.* Canberra: Australian National University. 112–171.

Skousen, Royal (1989). *Analogical modeling of language.* Dordrecht: Kluwer Academic.

Steriade, Donca (2000). Paradigm uniformity and the phonetics-phonology boundary. In Michael Broe & Janet Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University Press. 313–334.

Stump, Gregory (2001). *Inflectional Morphology.* Cambridge: Cambridge University Press.

Stump, Gregory & Raphael A. Finkel (2013). Morphological Typology: From Word to Paradigm. Cambridge: Cambridge University Press.

Szalai, Andrea (2007). Egységesség? Változatosság? A cigány kisebbség és a nyelvi sokféleség. In Bartha (ed.) (2007). 20–51.

— (2015). A cigány kisebbség nyelvei: szociolingvisztikai aspektusok. In Anna Orsós (ed.) *A romológia alapjai.* Pécs: PTE BTK Romológia és Nevelésszociológia Tanszék. 117–150.

Szilágyi N., Sándor (2011). Szinkrónia és diakrónia – de miről is beszélünk? In Szilágyi & Kádár (eds.) (2011). 9–30.

Szilágyi N., Sándor & Edit Kádár (eds.) (2011). *Szinkronikus nyelvleírás és diakrónia.* Kolozsvár: Erdélyi Múzeum-Egyesület.

Tálos, Endre (1988). Előszó. In Choli-Daróczi & Feyér (1988). 5–6.

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

Taylor, John R. (2003). *Linguistic Categorization.* Oxford: Oxford University Press.

Thomason, Sarah G. & Terrence Kaufman. (1988). *Language Contact, Creolization and Genetic Linguistics.* Berkeley, CA: University of California Press.

Tomasello, Michael (2009). The usage-based theory of language acquisition. In Edith L. Bavin (ed.) *Child Language.* Cambridge: Cambridge University Press. 69–88.

Törkenczy, Miklós, Péter Szigetvári & Péter Rebrus (2011). Harmony that cannot be represented. In Johan Brandtler, Valéria Molnár & Christer Platzack (eds.) *Approaches to Hungarian Volume 13 – Papers from the 2011 Lund conference.* Amsterdam: John Benjamins. 229–252.

Trager, George L. (1958). Paralanguage: A first approximation. *Studies in Linguistics* **13**. 1–12.

Turner, Ralph L. (1926). The position of Romani in Indo-Aryan. *Journal of the Gypsy Lore Society, third series* **6**. 129–138.

— (1927). The position of Romani in Indo-Aryan: a reply to Dr J. Sampson. *Journal of the Gypsy Lore Society, third series* **11**. 115–120.

Valis, Éva (1968). Two Gypsy tales from Hungary. *Acta Linguistica Academiae Scientiarum Hungaricae* **18:3–4**. 375–392.

Vekerdi, József (1961). Gypsy texts. *Acta Orientalia Academiae Scientiarum Hungaricae* **13**. 305–323.

— (1966). Mönch Bleihaupt (Zigeunermärchen). *Acta Linguistica Academiae Scientiarum Hungaricae* **16**. 135–148.

— (1971a). Statistisches zum Wortschatz des Zigeunerischen. *Acta Linguistica Academiae Scientarum Hungaricae* **21**. 129–134.

— (1971b). The Gurvari Gypsy Dialect in Hungary. *Acta Orientalia Academiae Scientiarum Hungaricae* **24**. 381–389.

— (1981). *A magyar cigány nyelvjárás nyelvtana. (Tanulmányok a cigánygyerekek oktatásával-nevelésével foglalkozó munkacsoport vizsgálataiból IV).* Pécs: JPTE Tanárképző Kar.

— (1982). *A magyarországi cigánykutatások története.* Debrecen: Kossuth Lajos Tudományegyetem.

— (1984). The Vend Gypsy dialect in Hungary. *Acta Linguistica Academiae Scientarum Hungaricae* **34**. 65–86.

— (1985). *Cigány nyelvjárási népmesék.* Debrecen: Kossuth Lajos Tudományegyetem.

— (2000). *A Comparative Dictionary of Gypsy Dialects in Hungary.* Budapest: Terebess.

Vekerdi, József & György Mészáros (1980). *A magyarországi oláh cigány nyelvjárás mondattana.* Budapest: Magyar Nyelvtudományi Társaság.

Vosniadou, Stella & Andrew Ortony (eds.) (1989). *Similarity and analogical reasoning.*

Cambridge: Cambridge University Press.

Wedel, Andrew (2007). Feedback and regularity in the lexicon. *Phonology* **24**. 147–85.

— (2009). Resolving pattern conflict: Variation and selection in phonology and morphology. In Blevins & Blevins (eds.) (2009). 83–100.

Wheeler, Benjamin I. (1887). *Analogy and the Scope of Its Application in Language.* Cambridge, MA: University of Cambridge.

Wilder, Chris, Hans-Martin Gärtner & Manfred Bierwisch (eds.) (1997). *The Role of Economy Principles in Linguistic Theory*. Berlin: Akademie Verlag.

Wolf. Siegmund A. (1960). Zur Frage einer normierten Zigeunersprache (Basic Romani). *Phonetica* **5:3–4**. 204–209.

Wright, Sue (2004). *Language Policy and Language Planning. From Nationalism to Globalisation.* London: Palgrave Macmillan.

Wurzel, Wolfgang U. (1989). *Inflectional Morphology and Naturalness.* Dordrect: Kluwer Academic.