

Report on documenting the Gurbet Romani variety in East Serbia and creating the Romani-Serbian dictionary



Mirjana Mirić – Svetlana Ćirković¹

The Romani-Serbian Dictionary of the Knjaževac Gurbet Variety (Ćirković & Mirić 2017) originated from work on the project *Exploring the Language and Folklore of Roma in Knjaževac*, financed by the Ministry of Culture and Information of the Republic of Serbia in two project cycles, 2016 and 2017. The project was spearheaded by the “Njegoš” National Library in Knjaževac. The linguistic team in cooperation with local Roma activists carried out field documentation of the language and folklore tradition of the Roma in Knjaževac and Minićevo.² The primary goal was the documentation of Romani language and tradition, the secondary the development and practical implementation of the specific field work methodology on a language insufficiently mastered by the researchers. The language material collected was used to compile a Romani-Serbian dictionary with a brief collection of selected texts in Romani. An important result of the project was the formation of a competent team of young Roma researchers, who today can successfully apply the knowledge acquired in field research and language data processing to their work with other Roma communities in Serbia.

Work on *Exploring the Language and Folklore of Roma in Knjaževac* went ahead in several stages in both project cycles. In the course of 2016, local Roma activists took

-
- 1 The authors are researchers in the project “Language, Folklore and Migrations on the Balkans”, conducted by the Institute for Balkan Studies of the Serbian Academy of Sciences and Arts in Belgrade (Serbia), financed by the Ministry of Education, Science and Technological Development of the Republic of Serbia.
 - 2 Knjaževac is a town in east Serbia, the village of Minićevo is 15 kilometres away. The Roma community is numerous in both town and village. Roma used to live in other villages in the vicinity, but due to social and economic factors they migrated to Knjaževac or other larger towns in Serbia. Many Roma from Knjaževac and the surroundings are temporarily working in Western Europe. The town of Knjaževac and neighbouring Minićevo were selected as sites for research in order to reconstruct the lexical material from the area, gathered in 1873 by the eminent Vienna philologist Franc Miklošič, and published in the sixth volume of his study *Über die Mundarten und Wanderungen der Zigeuner Europas*, entitled *Vocabularien der Mundart der Zigeuner in Serbien* (Mikloshich 1872–1880). It was unclear which variety of the Romani language that was. Reconstruction has established that the lexemes from Miklošič’s list belong to the Arli variety, which used to be spoken by Roma Blacksmiths in the town of Knjaževac and surroundings. They, however, in the course of time abandoned Romani and shifted entirely to Serbian (Sikimić 2017).



part in workshops providing training in field research, transcription of material and writing up the work. In addition to the training, field research took place the same year, producing about 14 hours of audio and video material.³ In the next project cycle (2017), training covered the lexicographical processing of the documented language material — excerpting of transcripts, selection of lexemes, lexicographical processing of selected words and translation into Serbian.⁴

Research on the ground showed three Roma groups at present living in Knjaževac and Minićevo: the Gurbets, Leyash and Blacksmiths, along with a few settled families of Arli and Bayash. The Gurbets in Knjaževac and Minićevo (who also call themselves *Zavradžije*) use the Gurbet variety, which belongs to the Southern Vlax dialect of Romani (dialectological division according to Yaron Matras, cf. Matras 2004: 7). The Leyash use a variety belonging to the Northern Vlax dialect of Romani (Matras 2004: 8, Boretzky 2003). The Blacksmiths today speak only Serbian, having lost their Romani in the mid 20th century, and are considered to be the descendants of the Arli, who speak a Balkan dialect of Romani (Matras 2004: 7, Sikimić 2017). Since the Gurbets are the most numerous group in Knjaževac and Minićevo, the dictionary is based on the Gurbet Romani variety.

In Serbia, there are no dialect dictionaries of local Romani varieties resulting from linguistic research in the field. Existing dictionaries have for the most part emerged from processing the author's lexicon (cf. Haliti 2011, Krasnići 2012) and contain lexemes not only from Gurbet but also from the Arli variety. The dictionaries, compiled by the same linguistic methodology used for gathering material and lexicographical processing, are mainly based on lexical material from different Romani dialects and varieties (cf. Boretzky, Iгла 1994). Rade Uhlik's dictionary (1983) is based on "the lexical fund of the Gurbet variety of Serbian Roma", together with other data from Romani varieties of the former Yugoslavia to which he had access (Uhlik 1983: 11). It is difficult, however, to determine to which particular dialect of the Romani the lexemes in it belong.

CORPUS

The corpus for the dictionary consists of transcripts of interviews recorded with native speakers of the Gurbet variety in Knjaževac and Minićevo in July 2016. A sample

3 The participants in the workshops, field research and transcription of documented material were Darko Ibrić, Katarina Ibrić, Miljan Mustafić, Željko Mustafić, Ružica Rakić, Ivan Ronić, Ljubinka R. Simić, Ljubinka S. Simić, Marija Simić, Romana Stefanović. Lectures were held by Biljana Sikimić and Svetlana Ćirković from the SASA Institute for Balkan Studies. The complete field material was deposited in the Digital Archive of the Institute for Balkan Studies in Belgrade and in the "Njegoš" National library in Knjaževac.

4 In the project cycle of 2017 there were fewer Roma associates: Darko Ibrić, Katarina Ibrić, Ljubinka R. Simić, Ljubinka S. Simić and Zagorka Simić; lectures were held by Biljana Sikimić, Mirjana Mirić and Svetlana Ćirković. In addition, the following native speakers of the Gurbet Romani provided their help in the overall process of creating the Dictionary and editing: Darko Ibrić, Katarina Ibrić, Ljubinka S. Simić (henceforth: editors).



was gathered from respondents of both genders and varying age, born in Knjaževac and the surrounding area or living in Knjaževac for an extended period, which makes them representative speakers of the local Gurbet variety. We should note that all respondents are bilingual in Romani and Serbian, while many also speak German, having lived for lengthy periods in Germany.

The recorded material amounted to about 14 hours and contains Leyash and Arli in addition to Gurbet, as well as Serbian. Regarding transcription, apart from the segments in Serbian, we also transcribed fragments of interviews in Leyash and Arli, but for linguistic reasons it was decided not to include this in the final product. The dictionary, therefore, is based on 7 hours and 30 minutes of exclusively Gurbet variety — over 45,000 words. The lexemes were excerpted from 12 transcripts of interviews and marked by numbers and square brackets, denoting transcripts of Gurbet variety: [Tr2], [Tr4], [Tr5], [Tr6], [Tr8], [Tr9], [Tr11/1], [Tr11/2], [Tr11/3], [Tr11/4], [Tr16], [Tr17]. Recorded interviews with the Leyash and transcribed interview fragments are marked: [Tr7], [Tr13], [Tr15]. There is only one recorded interview in Arli variety ([Tr1]), and three predominantly in Serbian: [Tr3], [Tr12], [Tr14].⁵

The method used was the open interview: questions were asked in Romani by native speakers, in Serbian by associates or researchers whose first language was not Romani. They related to traditional culture, since the goal of the project was to gather language and folklore material, elements of traditional culture still living in the traditional practice of the local Roma community, but also elements which could only be reconstructed from the respondents' memories. The interviews also covered topics from oral history and (auto)biographical stories. This methodology resulted in spontaneous production of the registered lexical fund in narratives; no special questionnaires were used to elicit lexemes and their grammatical forms. The lexemes reflect the topic of conversation. At the end of the book there are 12 samples of idiolects reflecting the realistic and current use of the lexemes given in the dictionary. Each sample gives the transcript number together with the origin and gender of the respondent. For ethical reasons, the names of the respondents have not been given. The texts have been chosen so as to give a broader context for the use of words that found their way into the dictionary, and to illustrate elements of traditional culture, the past and present life of the respondent.⁶ Due to limited space,⁷ the selected texts are short, one fragment being chosen for each respondent.

5 Although in the Gurbet variety, the interview and transcripts marked [Tr10] were not used due to technical problems.

6 A more extensive selection of texts in Romani with Serbian translation will be found in a monograph to be published later.

7 For technical reasons, the publisher defined the number of pages before work on the dictionary was completed, which resulted in a decision to include, in addition to Romani lexemes, only the most frequent borrowings from Serbian adapted into the Romani language system. However, the dictionary could have included non-adapted loanwords which would have demonstrated the entire lexical system in current use.



SPELLING AND ALPHABET

The spelling used in the dictionary follows the model recommended by Rajko Đurić (Đurić 2012), with some slight deviations to reflect the dialectological transcription of the material. As such it is non-standard, but rather serves to mark the dialectological characteristics of local oral speech.

The following graphemes were used in the dictionary: a, b, c, č, čh, ć, ćh, d, dj, đ (only in the Serbian, non-adapted words), dž, e, ə, f, g, h, x, i, j, k, kh, l, lj, m, n, nj, o, p, ph, r, rr, s, š, t, th, u, v, z, ž.

LEXICAL CHOICE

For dictionary requirements, all words from Romani were taken from the interview transcripts; however, also included are some obviously recent loanwords from Serbian adapted into the Romani morphological system. The dictionary also contains terms that denote the elements of material and intangible traditional culture, whose occurrence in the interviews is the result of steering the conversations towards topics of traditional culture. Besides these terms (e.g. *Bibijako djive* 'Aunt's day'), we also recorded some that came from Serbian (e.g. *Božić* 'Christmas').

It is important to note that the presence or absence of words from the lexicon of some idiom may be an important indicator of the different social, historical or cultural circumstances in which language develops. The material points to a tendency to lose certain Romani words and an inevitable borrowing from Serbian, as the dominant language of a broader community. The loanwords have become the integral part of the lexicon of the speakers and have mostly been adapted into the Romani morphological system; the way in which they are adapted reflects vital language processes.

Though relevant, certain excerpted and processed words did not make it into the final version of the dictionary due to restricted space. These are nouns, adjectives, verbs and adverbs (and an insignificant number of prepositions) originally from Serbian (more rarely from other languages, such as German) whose use in the material is quite rare, usually registered for only one speaker and often a hapax. Some originally Serbian words have their equivalent in Romani (e.g. *učil* — *sićol* 'learn'), while for some there is no confirmed Romani alternative either in the transcripts or known to the editors (e.g. *mrzol* 'hate').

The following groups of words have been omitted from the dictionary: the entire onomastic system, recent borrowings from German and most nouns borrowed from Serbian whose meaning remains transparent, despite their morphological adaptation into the Romani language system. Exceptionally, lexemes originating from Serbian and used by only one respondent have been included. They have the value of a term in view of the fact that the topics of the interviews also related old crafts and the memories of older technological procedures.

The dictionary does not contain etymological information (those interested in etymology have numerous relevant sources at their disposal). It is important, however, to note that apart from autochthonous words, pre-European and European loanwords

characteristic of Vlax Romani varieties (cf. Boretzky, Igla 1994 311–338) have been included. Also included are older loanwords from Slavic languages (e.g. *čeljado* ‘a child’), of which some may have entered the Gurbet variety through Romanian, as well as older Turkish words rarely in use today (e.g. *bočava* ‘an apron’). In such entries the Serbian equivalent is always given as part of a series of translation synonyms.



ENTRIES

The dictionary numbers 920 entries (without references or expressions). In forming an entry, the accented lemma is given first in Romani, then the preterite forms (of verbs), the plural (of nouns), the part of speech with its grammatical categories, and finally the meaning is translated into Serbian.

The words are transcribed from the audio-material just as the respondents spoke them. There is a notable variation in the pronunciation of phonemes not only among respondents but even with the same respondent. When it became clear that the error was an accident of pronunciation, in agreement with the editors certain words were corrected in the entry while the authentic pronunciation was recorded in the examples.

If a word appears in more than one meaning in the material, then it is separately cited in the entry, with additional meanings that were not registered in the material but which the editors singled out as relevant for the dictionary. Each word is used in the context that best illustrates how it may actually be used, thus enabling verifiability of the meaning and lexicographical procedure. The context may also serve for elementary linguistic analysis.

Each entry is illustrated by examples from the transcripts, and each example carries the mark of the transcript from which the word was excerpted.

LEXICOGRAPHICAL PROCESSING OF WORDS

In the dictionary, all two-syllable and multi-syllable words (as the main feature of the entry with expressions given separately) are marked with accents which can be verified from the recorded material. Examples of use are not accented since this would hamper reading. We should underline that accents were additionally checked and corrected by the editors. Except in rare cases (cf. Boretzky, Igla 1994), authors of existing Romani dictionaries for the most part do not provide the accent (cf. Haliti 2011, Kajtazi 2008, Krasnići 2012, Uhlík 1983, Romlex). We consider that the inclusion of accents is invaluable, not only for dictionary users to learn the pronunciation of a given word, but also for linguists who can follow the manner in which forms of plural nouns or the preterite of a verb are accented, or how the accent of loanwords is adapted and the like.

The form of the entry and grammatical categories cited depend on the part of speech.

Nouns are mainly given in the nominative singular, while for plurale tantum nouns it is pointed out that this is a form of the plural. The nominative plural is also



reconstructed and the gender is given (masculine or feminine). Reconstruction of the plural form of nouns is based on the transcripts; however, if the material does not confirm the plural form, it is reconstructed by the editors. For certain concrete uncountables (e.g. *arro* 'flour'), or abstract nouns (e.g. *bax* 'fortune') the plural form is not given since it has not been confirmed either in the material or by the editors.

Adjectives are given in the masculine and feminine gender singular and plural, while the translation into Serbian contains only the masculine singular (e.g. *baro*, *bari*, *bare* adj. 'big'). The same applies to adjectival pronouns which have forms in masculine and feminine gender and singular and plural number (e.g. *akava*, *akaja*, *akala* dem. pron. 'this'), while personal pronouns are given in the nominative singular as their basic form, and all other registered case forms are given as separate entries which refer to the main entry (e.g. *manca* pron. cf. *me* 'I').

The numerical category includes cardinal and ordinal numbers as well as certain words which, for the requirements of this dictionary, have been marked as a number. Viewed linguistically, it is possible to classify them differently (e.g. the number *milja*, 'thousand' — may be treated as a noun).

Since the articles do not exist in the Serbian language, Romani articles were not translated. It is important to point out that in the system of articles described in the existing literature which indicates the presence of the articles *o*, *i*, *e* (cf. among others, Đurić 2005, Leggio 2011), neither the material nor the editors found confirmation for the article *i*. The article *o* is used with masculine nouns in the singular, while the article *e* is used with feminine nouns in the nominative singular, nouns of both gender in the nominative plural, and also with nouns of both gender and number used in the oblique case.

Verbs are given in the form of the third person singular, present tense, which is standard in current Romani lexicography (cf. for instance Boretzky, Iglá 1994, Romlex, Kajtazi 2008, Haliti 2011, Krasnići 2012). Reflexive verbs are given with the reflexive pronoun *pe*, but some verbs can also have a non-reflexive meaning. In such cases the reflexive pronoun *pe* is placed in a bracket, e.g. *iril (pe)*, *žalil (pe)*, *ženil (pe)*, and the meanings are given separately for the reflexive and non-reflexive form. The third person singular of the past tense is also given in brackets, another lexicographical standard. As the examples from the corpus do not always confirm the present or past tense, the missing forms were reconstructed with the help of an editor. The use of all forms of verbs found in the interview transcripts is illustrated by examples. Besides verbs, expressions are also given with examples, e.g. *del andre*, *xal sovli*. Since in some expressions a noun has been used with the verb, the noun lexeme is referenced to the verb.

Although in the system of parts of speech, the copula does not figure as a separate part but as a syntactical function of the verb 'to be', in the dictionary the verb 'to be' has been marked grammatically as a copula. We decided on this in accordance with the present practice of Romani lexicography (cf. for example Boretzky, Iglá 1994: 258, Romlex s.v. *si*).

In the lexicographical processing of adverbs, the lexeme is given in the form in which it predominantly appears. Apart from monolexemic adverbs, adverbial phrases are also cited. These may be Romani in origin, e.g. *thara (o) djive* 'tomorrow,

the next day', *de kana* 'since', or a combination of a Romani and a Serbian word, e.g. *preko drom* 'across', *kozom god* 'whatever (the quantity)'. Also included in the dictionary are adverbs that originally came from Serbian, but the frequency of their use and the possible non-use of a Romani word of equivalent meaning signals that they are becoming part of the lexicon of the local Gurbet Romani variety and as such ought to be included, e.g. *obavezno* 'necessarily', *tek* 'just'.

Being an unchangeable and functional part of speech, prepositions are grammatically marked in the dictionary. Since some have two variants depending on the article of the noun they accompany, e.g. *ande, ando* 'in, into'; *pe, po* 'on', both forms are given separated by a comma, but the form with an article for the masculine gender, e.g. *ando, po* is set apart as an entry referring to the main entry, e.g. *ande, pe*. The reason, both in the dictionary and the reference, is the alphabetical order. The prepositions borrowed from Serbian are *za* 'for' and *sa* 'with', confirmed in a large number of examples from the corpus.

Almost all conjunctions are borrowed from Serbian, exceptions being *ke* 'because' and *thaj* 'and', with the variant *the*, but this conjunction also has a semantic variant borrowed from Serbian, the conjunction *i* 'and', confirmed in the corpus by a large number of examples.

Interjections as functional words are not confirmed in large numbers. We have abstracted *ake* ('evo', 'here'), *eke* ('eto' 'there').

Particles in the Romani language are represented by words with a morpho-syntactic function — they are used to form the analytic future — the particle *ka*, for negation — the particle *ma, na, ni*, for comparison — the particle *maj*, for the conjunctive — the particle *te*. The particle *te* is in some examples used as a conjunction meaning 'if', and these meanings are given separately. We should single out the particle *naj* which is used as the negation form of the third person singular of the copula 'to be' in the present tense and the negation form of the verb 'to have' in the present tense.

THE USAGE VALUE OF THE DICTIONARY

The main motivation in making the dictionary was to illustrate the local Gurbet variety through a selection of entries, examples and spoken models. It can be of use and interest to the exponents themselves, particularly to younger speakers who at the crucial stages of language acquisition, are increasingly less exposed to Romani and are shifting to Serbian, or evidence features of an asymmetric bilingualism. To those knowledgeable in other varieties and dialects of Romani, the dictionary can be interesting for comparative purposes, and to linguists for research, since it provides complete morphological information on each entry as well as giving a context for use of the word: a source of material for linguistic research at various levels of linguistic structure. For anthropologists and ethnologists, the use of particular ethnographic terms can indicate traditional culture, the extent to which it is present and valuable in daily life. Ultimately, the dictionary is intended for all who are interested in learning Romani.





Since the dictionary was created based on a sample of oral speech, we should point out that it is not intended to be prescriptive. It is characteristic of spontaneous oral speech that it deviates from other forms, especially from the written language. Even though the examples given in the entries may appear incomprehensible or be considered ‘wrong’ or ‘corrupted’, they are not. They demonstrate the typical features of oral speech at different levels of language structure, such as variation in speed of the utterance, omitting words, frequent reference to context, false starts and corrections, the accompanying elements of non-verbal communication which visually complement the statement; differences in syntactic structure are especially perceptible.

The samples of idiolects quoted at the end of the book point to differences in individual use of Romani with different respondents, depending on their language biography, and also the use of different groups of words in accordance with the topic of conversation.

The UNESCO Atlas of World Languages in Danger registers Romani in 28 European countries and classifies it as ‘definitely endangered’. In 2006, Serbia ratified the the European Charter for Regional or Minority Languages (ECRML) where Romani was given minority language status.

Field documentation of the language and folklore material in the *Exploring the Language and Folklore of Roma in Knjaževac* project complements published studies of Romani by major West European universities (Manchester, Graz, Prague etc.). At local level in Serbia, this project can serve as a model for documenting the local variety of Romani and the protection of endangered languages. In documenting folklore traditions and elements of traditional culture, a model has also been provided for selecting those elements of the intangible cultural heritage of the Roma that merit the protection of cultural institutions in the Republic of Serbia.

LITERATURE

- Boretzky, N. — Iglá, B. 1994. *Wörterbuch Romani — Deutsch — Englisch für den südosteuropäischen Raum*. Wiesbaden: Harrassowitz Verlag.
- Boretzky, N. 2003. *Die Vlach-Dialekte des Romani. Strukturen — Sprachgeschichte — Verwandtschaftsverhältnisse — Dialektkarten*. Wiesbaden: Harrassowitz Verlag.
- Ćirković, S. — Mirić, M. 2017. *Romsko-srpski rečnik knjaževačkog gurbetskog govora*. Knjaževac: Narodna biblioteka “Njegoš”.
- Đurić, R. 2005. *Gramatika romskog jezika*. Beograd: Otkrovenje.
- Đurić, R. 2012. *Standardizacija romskog jezika*. Sarajevo: Udruženje „Kali Sara”, Romski informativni centar.
- Haliti, B. 2011. *Srpsko-romski rečnik sa gramatikom i pravopisnim savetnikom*. Novi Sad: Prometej.
- Kajtazi, V. 2008. *Romano-kroacijako thaj kroacijako-romano alvari*. Zagreb: Odjel za orijentalistiku Hrvatskog filološkog društva i „Kali Sara”, Udruga za promicanje obrazovanja Roma u Republici Hrvatskoj.
- Krasnići, A. 2012. *Angluno rromano-srbikano, srbikano-rromano alvari*. Subotica: Rromane pustika.
- Leggio, D. V. 2011. The dialect of the Mitrovica Roma. *Romani Studies* 5: 21(1), 57–114.
- Matras, Y. 2004. *Romani. A Linguistic Introduction*. Cambridge University Press.

Mikloshich, F. 1872–1880. *Über die Mundarten und Wanderungen der Zigeuner Europas I–XII*. Wien: Karl Gerold's Sohn.
 Romlex Lexical Database — available at: romani.uni-graz.at/romlex/lex.xml

Sikimić, B. 2017. *Mikošičeva zbirka reči „Iz Timoka”*. Knjaževac: Narodna biblioteka „Njegoš”.
 Uhlik, R. 1983. *Srpskohrvatsko-romsko-engleski rečnik*. Sarajevo: „Svjetlost”, OOUR Izdavačka djelatnost.



Mirjana Mirić is a research associate at the Institute for Balkan Studies of the Serbian Academy of Sciences and Arts (Belgrade, Serbia), engaged on the project *Language, Folklore and Migrations in the Balkans*. She holds a B.A. degree in Linguistics and MPhil degree in Language Acquisition (University of Belgrade). She completed the Specialization in Linguistics (University of Geneva) in 2010, and defended her doctoral thesis in 2016 at the University of Niš, Serbia. In 2017 she collaborated on the project *Exploring the Language and Folklore of Roma in Knjaževac* (The “Njegoš” National library, Knjaževac, Serbia). Her research interests include Balkan linguistics, Romani linguistics, bilingualism, minority and endangered languages, first and second language acquisition, and corpus linguistics.
 Contact: mandic.mirjana@gmail.com

Svetlana Ćirković is a senior research associate at the Institute for Balkan Studies of the Serbian Academy of Sciences and Arts (Belgrade, Serbia), engaged on the project *Language, Folklore and Migration in the Balkans*. She holds a B.A. and MPhil degree in Serbian Language and Literature. She defended her doctoral thesis in 2011 at the University of Belgrade (Serbia). She was engaged on numerous national and international projects, focused on documenting various languages and language idioms in the Balkans. In 2016 and 2017 she collaborated on the project *Exploring the Language and Folklore of Roma in Knjaževac* (The “Njegoš” National library, Knjaževac, Serbia). Her research interests include Balkan linguistics, Romani linguistics, endangered languages, language documentation, and linguistic anthropology.
 Contact: scirkovic@hotmail.com