Romani

Yaron Matras

(University of Manchester)
2011

http://romani.humanities.manchester.ac.uk/virtuallibrary

EUROPEAN COMMISSION
European Research Area

SEVENTH FRAMEWORK PROGRAMME

Funded under Socio-economic Sciences & Humanities

Romani

Yaron Matras

1.   Introduction

Romani is the only Indo-Aryan languages that has been spoken exclusively in Europe (and by European emigrants to the New World) since the Middle Ages. It was apparently brought into Anatolia under Byzantine rule sometime around the eleventh century by a caste-like population specialising in itinerant services such as metalwork and entertainment. Their self-appellation, *řom*, appears to be related to that of the *ḍom* – a low-status and landless caste specialising in manufacturing small tools and artefacts, weaving, and entertainment. Other groups with cognate names, a similar socio-economic profile and an Indo-Aryan language living outside of India include the *dom* of the Middle East and the *lom* of the Caucasus. The *ḍom* are often described not just as segregated but also as an enslaved population, and it has been suggested that the ancestors of the *řom* may have been camp-followers who accompanied the Seljuk invasions through western Asia and were either captured or recruited by the Byzantines. Historical sources testify to a wave of Romani migrations from the Balkans into western and northern parts of Europe during the fifteenth century. The names most commonly used during this period to refer to them by the majority sedentary population are 'Gypsy', 'Yifti', 'Gitan' etc., from 'Egyptian' (based on a misguided assumption about their origin), and 'Tsigan', 'Zigeuner', 'Tsigngani', 'Çingene' etc., possibly from an old Turkic appellation for low-caste slaves (*čiɣan*).

Settlement in the various regions of Europe led to the dispersal of Romani populations and to their isolation from one another, and this, along with the influence of different contact languages, is what gave rise to the emergence of considerable differences among the dialects of Romani. The earliest known attestation of Romani has only just recently been discovered in a manuscript dating from around 1515 composed by by Johannes ex Grafing, who may have collected his material in Vienna (Knauer 2010). A text

published by Andre Borde in 1542 contains sample phrases in Romani thought to have been collected in England or France. The earliest known attestation from the Balkans is in the travel diary of Evliya Çelebi from 1668 (Friedman & Dankoff 1991). Sources from the mid-eighteenth century offer a sufficiently wide coverage to be able to establish that differences among the dialects of Romani by that time closely resembled the patterns of differentiation found today. It was in this period that hypotheses concerning the Asian origin of Romani were first made, culminating in Johann Rüdiger's (1782) publication that conclusively proved that Romani was closely related to Hindustani. Today, Romani constitutes one of the largest minority languages of Europe, with upwards of 3.5 million speakers residing mainly in southeastern Europe (Romania, Bulgaria, Serbia, Macedonia, Greece, Albania) as well as in central Europe (Slovakia, Hungary). There are also sizeable Romani-speaking communities in the United States and in South America as well as in urban centres of western Europe. In most of the fringe regions of Europe – Scandinavia, Britain, and the Iberian Peninsula – Romani populations abandoned their language toward the beginning of the nineteenth century or even earlier, but maintain an in-group vocabulary of Romani origin, which comprises up to 600 lexical roots. All Romani speakers are bilingual and often multi-lingual from a very young age. The language has traditionally been limited to oral usage, primarily within the extended family and with neighbouring clans. Code-switching and mixing are common and dialects of Romani have absorbed considerable structural and lexical influences from their respective contact languages.

The lexicon of most Romani dialects contains a shared inventory of only around 1,000-1,200 lexical roots. Of those, only around 800 are pre-European, consisting of a core vocabulary of around 600 Indo-Aryan roots along with layers of early loans from Iranian languages and from Armenian (as well as other Caucasian languages). All Romani dialects also contain a significant layer of Greek loans, as well as a much smaller number of shared lexical items of Balkan Slavic and Balkan Romance origin. The remainder of the vocabulary is typically derived from subsequent contact languages, and thus differs among the individual dialects of the language. The shared lexicon provides some additional clues to the history of the population, suggesting a prolonged settlement in an area in which Greek, Armenian, and Iranian will have played a significant role as contact languages – in all likelihood eastern or central Anatolia – and a shared migration into the Balkans before the dispersal into groups.

Traditionally an oral language of the household, Romani is increasingly being used mainly in electronic communication among community activists, who practice a pluralistic approach to codification and the selection of dialect features. Some programmes to promote the language in education and media are underway, though there is no standard written form of the language nor a prestige variety that would led itself for the creation of such standard.

2.   Phonetics and phonology

2.1.   Phoneme inventory

Romani has five inherited phonemic vowel qualities /i/, /e/, /a/, /o/, /u/. Many Romani dialects also have a central or centralised vowel, /ə/ or /ɨ/ or sometimes /ɯ/, which may be limited to loanwords but is often found to infiltrate the inherited lexicon as well. Some dialects take on additional phonemes, such as /y/ (in contact with Turkish) and /ɐ/ (in contact with German and English). In contact with western and central European languages, some Romani dialects acquire vowel length distinctions, which, however, are not always phonemic. There are few inherited words containing the diphthongs /aj/,/oj/, and /ej/.

The consonantal phonemes are (i) plosives, (a) voiceless, non-aspirated: /p/, /t/, /k/, (b) voiceless, aspirated /pʰ/, /tʰ/, /kʰ/, (c) voiced: /b/, /d/, /g/; (ii) affricates, (a) voiceless, non-aspirated: /ts/, /tʃ/, (b) voiceless, aspirated: /tʃʰ/, (c) voiced: /dʒ/; (iii) fricatives, (a) voiceless: /f/, /s/, /ʃ/, /χ/, /h/, (b) voiced: /v/, /z/; (iv) nasals: /m/, /n/; (v) liquids, (a) lateral: /l/, (b) rhotic: /r/, (vi) approximant: /j/. A marginal phoneme is /ʒ/ is often limited to loanword. Many Romani dialects show palatalisation of consonants in positions preceding front vowels, and some have palatal phonemes such as /c/ and /ɟ/. Conservative Romani dialects in southeastern Europe tend to preserve two rhotics, a trill /r/ and an additional sound that is realised either as a uvular /ʀ/ or even a voiced velar fricative /ɣ/, or in some dialects as a retroflex /ɽ/. Some dialects show a collapse of /χ/ and /h/ in /x/.

2.2.   Stress

Conservative Romani stress patterns are grammatical. A series of older inflection suffixes carry stress, e.g. the nominative ending in *rakl-ó* 'boy', *rakl-í* 'girl', and the person ending in *li-jóm* 'I took', *dikh-jó-l* 'it is seen'. A secondary or historically younger set of grammatical suffixes are external and do not carry stress, e.g. the remoteness tense marker *-as* as in *dikh-l-jóm-as* 'I had seen', or so-called Layer II case-markers that following the oblique ending, as in *rakl-és-ke* 'for the boy'. In some, especially the western dialect of Romani, stress patterns are taken over or partially replicated from the contact languages. Stress shifts forward and becomes fixed to a particular syllable rather a grammatical ending.

2.3.   Syllable structure, phonotactics, morpho-phonology

The typical syllable structure is CV or CVC, with most words having up to three or at the most four syllables, typically CVCVCV or CVCCV(CV), etc. Consonant clusters in the inherited lexicon are largely limited to the combinations Cr (in initial or medial position) and Cn, Cm (mainly in medial or final position), though through the large number of

loanwords from different languages syllable patterns show enormous variation across the dialects. Several issues involving morpho-phonology are noteworthy. First, there is a tendency for segments representing person endings on the past-tense verb as well as medio-passive and sometimes also causative derivational morphemes to show jotation, which in turn often leads to palatalisation of the preceding segment: Thus *gel-jom* 'I came' > *gejom* (but *gel-o* 'he came', with an adjectival-participial ending), *kerd-jov-el* 'it is being done' > *kerd'ovel, kerdžovel.* Next, Romani as a whole shows alternation in grammatical paradigms between the phonemes /s/ and /h/, with individual dialects opting for different solutions. The most likely position for /h/ to appear is in morphological endings in intervocalic position, thus Russian Romani *le-sa* 'with him', *džin-es-a* 'you know', German Romani *le-ha* 'with him', *džin-eh-a* 'you know'. Alternations are also found in the root form of the copula verb: Erli Romani from the southern Balkans *sin-* 'is', Finnish Romani *hin* 'is'. Some dialects show /h/ in the interrogative paradigm: East Slovak Romani *sar* 'how', West Slovak Romani *har* 'how'. Third, Romani second-layer or external case-markers are enclitics that show voice assimilation to the preceding segment: *tu-tar* 'from you', *les-ke* 'for him', but *man-dar* 'from me', *len-ge* 'for them'. Finally, Romani past-tense or perfective markers show inflection class re-assignment based on a hierarchy of features represented by the final segment of the verbal root: *-m* > *-k, -kh* > *-č, -čh* > *-š, -s*, with stems at the top of the hierarchy more likely to be re-assigned from the class with in *-t-* to that in *-l-*: **mukh-t-jom* 'I left' > *mukh-l-jom* (see Matras 2002: 138-139).

3.   Morphology and grammatical categories

3.1.   Morphological type

Typologically, Romani morphology is partly fusional and partly agglutinative. The fusional element tends to be older and is best represented by the person conjugation on the verb, especially in the present tense, and by the retention of declension classes in the first layer of nominal inflection, thus: *rakl-o,* 'boy (nominative)', *rakl-e* 'boys (nominative)', *rakl-es* 'boy (oblique)', *rakl-en* 'boys (oblique)', *rakl-i* 'girl (nominative)', *rakl-jan* 'girls (oblique)', and so on. (Note that nominal inflection of this type also appears in personal pronouns, demonstratives, interrogatives, definite articles and adjectives). The agglutinative element is generally younger and is best represented by second-layer case-markers, which are enclitic and not sensitive to declension class: *rakles-tar* 'from the boy', *raklja-tar* 'from the girl'. Some tense and modality markers as well as verb derivational markers indicating transitivity and intransitivity are also agglutinative: Russian-Baltic Romani *sikhlja-kir-* 'to teach, to show', *xolja-kir-* 'to upset, to anger'. Some older derivational markers – for example the transitive/causative augments *-av-* and *-ar-* retain their agglutinative character (*dar-* 'to fear', *dar-av-* 'to frighten'; *dand* 'tooth', *dand-ar-* 'to bite'), while some younger markers are in

the process of losing their agglutinating-derivational character and are becoming fusional-inflectional: *ker-el* 'he/she does', *kerd-jov-el* 'it is done' > *kerdj-ol* 'it is done', and by analogy also *ačh-ol* 'he/she stays' and other intransitive or situational verbs. In addition to the pre-European morphological component, Romani dialects inherit an inventory of Greek-derived inflectional markers, such as the tense-aspect markers *-in-*, *-az-* and so on. In many dialects these markers take on the function of agglutinative derivational elements, identifying the root as a loan: thus Turkish Romani *anlad-iz-av* 'I understand' (from Turkish *anladı* 'understood', with Greek-derived *-iz-*), Ursari Romani (Romanian Black Sea coast) *entr-iz-av* 'I enter' (Romanian *într-* 'enter', with Greek-derived *-iz-*). Suppletion is rare in Romani and appears primarily in the pronominal system, where we find a split between renewed nominative forms in *ov* 'he', *oj* 'she' etc., and more conservative oblique forms in *les* 'him', *la* 'her' and so on (the same split also affecting the set of demonstratives and in some dialects also the definite articles), and in the verb 'to go' (present *dža-*, past *ge-*). Further suppletion often comes about through reliance on a loan word for individual words such as ordinals (especially 'first') and comparatives (especially 'worse'). On the whole the language can be said to be drifting toward stronger reliance on analytical structures. This is in part a contact phenomenon. The Romani dialects of western Europe have largely abandoned flexional derivations in the verb and rely instead on analytical constructions with auxiliaries. In southeastern Europe (Albania, Macedonia, and Greece), Romani dialects have developed analytical perfect tenses, while in central and eastern Europe medio-passive constructions tend to rely on the active verb with a reflexive pronoun. Inflectional case markers also show strong competition from prepositions, though the extent varies among the dialects. An outstanding characteristic feature of Romani morphology is the division between pre-European vocabulary and European loanwords in both nominal and verbal morphology. In nouns, the difference is expressed primarily in the choice of declension class, which determines the expression of the nominative and partly also the oblique form. In verbs, European loans require a loan-verb adaptation marker, which functions essentially as a derivational marker that mediates between the verb stem and its tense-aspect and person inflection.

3.2. Word structure

The majority of words inflect: nouns, adjectives, lower numerals, demonstratives, verbs, interrogatives, personal pronouns, definite and indefinite articles, local relations adverbs, and some indefinites. Words that do no inflect are conjunctions, interjections, and prepositions (though the latter are often derived from inflected local relations adverbs). The word typically consists of a word root, followed by a derivational marker and the by inflectional markers. Most morphological material is thus suffixed. Prefixed elements are found in the

noun phrase, and are usually analytical; they include attributes of various kinds – definite and indefinite articles, adjectives, prepositions, and genitive-possessive nouns. On the verb, the only regular prefixed structure are the aspectual markers borrowed from Slavic languages in the Romani dialects of Slovakia, Poland, Russia and the Baltic regions. Similar structures are emerging in the Romani dialects of Bulgaria, Macedonia as well as Greece and Austria, modelled on or directly borrowing aktionsart prefixes of the respective contact languages. Most grammatical morphemes have their fixed positions and there is a clear distinction in this respect between bound and unbound morphemes. A possible exception and a case for cliticisation is the set of more conservative nominative pronominal forms in *(l)o* etc.. In some dialects, they occur exclusively in non-verbal predications, where they tend to accompany demonstrative, interrogatives, or adverbs (e.g. Balkan Romani *kaj-lo?* 'where is he'), or in copula predications, while in other varieties they may also accompany lexical verbs and have rather flexible distribution pattern within the sentence (cf. German Romani *phenas-o miro phral ap mande-lo* 'my brother [he] said [he] to me'). Only one Romani dialect, that of Epiros in northwestern Greece (Matras 2004) is known to have incorporated oblique personal pronouns as bound morphemes into the verb: *dikhl-jom-os* 'I saw him', *dikhl-jom-i* 'I saw her'. Compounding is extremely rare.

3.3.   Nominal categories

There are two genders, masculine and feminine, which are always distinguished formally in the singular. The plural is also sensitive to gender distinctions, though not in personal pronouns or in adjectival-attributive inflection, and in many dialects not in the definite article or in oblique forms of the noun, either. Nominal declension paradigms are otherwise quite stable and are generally clearly distinguished for gender and number. Declensions are represented at the so-called Layer 1 level of nominal case. This is for most classes the level at which fusional case markers are direct descendents of Old and Middle Indo-Aryan case markers: The masculine singular nominative is expressed either by a vocalic ending *-o* or by the consonantal ending of the word root, the masculine singular oblique is expressed by *-es-* or by *-as-*, the feminine singular nominative is expressed by *-i* or by the consonantal word root, and the feminine singular oblique is expressed by *-a-*. Plural formation has *-e* in masculines ending in a vowel and *-a* in most other cases. In addition to the various declension classes for pre-European nouns, a set of declension classes exists that specialise in European loans (including most Greek loans). These vary somewhat among the dialects, both in the retention of differences and the specific shape of the morphs. The historical classes are masculines nominative in *-os*, *-us*, *-o* and *-is*, and feminine nominatives in *-a*. Oblique forms are generally subject to contamination between the European nominative and

the pre-European oblique forms, with a strong tendency toward analogies to the pre-European forms, while plural endings vary considerably, dialects often borrowing plural endings from contemporary contact languages. Layer 2 case markers consist of a fixed set of agglutinating markers that follow the fusional oblique stem of the noun. Cases that are distinguished at this level are the Dative (which encodes the beneficiary and recipient), and Locative (which encodes location, possessor, target, and is also used as a default prepositional case), the Ablative (source or substance), Sociative-Comitative (also used as Instrumental), and the Genitive. As in other Indo-Aryan languages, the Romani Genitive shows adjectival agreement with the head: *rakl-es-ker-o dad* 'the boy's father', *rakl-es-ker-i daj* 'the boy's mother'.

Personal pronouns show their own individual blique forms: *me* 'I', *man-* 'me'; *ov* 'he', *les-* 'him'. The regular Layer 2 markers then attach to the oblique pronominal stems: *man-sa* 'with me'. Possessive forms of the first and second person differ from those of the nominal paradigm, showing an old possessive formation in *-ro-* attached to the oblique stem: *miro* 'my', *tiro* 'your', *amaro* 'our' etc.

Animacy is of relevance to the marking of direct objects. Animate direct objects take the unmodified oblique case, which is often referred to as an 'accusative'. Inanimate direct objects take the nominative. The oblique also functions as the case of the possessor, in most dialects, and sometimes also of the external possessor, while in some dialects it may also accompany a preposition, indicating an indirect objects.

Definiteness is marked by preposed definite articles, which agree with the noun in gender, number, and case: *o graj našl-o* 'the horse ran away'. Indefiniteness may be marked or unmarked, depending on the topical status of the noun. Adjectives and demonstratives are generally preposed and agree with the noun in gender, number, and case: *mir-o bar-o phral* 'my big brother', *aka-ja phur-i džuvl-i* 'this old woman'.

Attributive agreement (as used in definite articles, demonstratives, and adjectives) is reduced to marking the gender of singular nouns while plurality is marked without distinguishing gender, and in the domain of case to marking the Layer 1 contrast between nominative and oblique. Only some dialects extend attribute agreement to the full set of Layer 2 case markers.

Romani shows an exceptionally complex, four-term system of demonstratives. The set encodes the immediate presence/absence of the referent (often understood as proximity/remoteness, but usually defined in terms of the location of the referent in the physical setting of the conversation or in the abstract realm of the discourse), as well as the degree of specificity or (dis-)ambiguity of the referent, thus *adava* 'this one (immediately visible and unambiguous)', *akava* 'this one (immediately visible, distinguished from other potential candidates)', *odova* 'that one (aforementioned and unambiguous)', *okova* 'that one (aforementioned, distinguished from other potential candidates)'. The forms themselves are subject to numerous analogies, re-shaping, and re-duplication processes.

3.4.   Verbal categories

Romani verbs are derived from verbal roots (*xa-* 'to eat'), from de-nominal stems (*dand-ar-* 'to bite', from *dand* 'tooth') or from de-adjectival stems (*bar-jov-* 'to grow', from *bar-o* 'big'). The verb stems consists of the lexical root, which may be followed by a derivation affix. The most common derivational affixes are the intransitive form *-jov-* (derived from the auxiliary *jov-*, from **bhuv-* 'to become') which is often contracted to *-jo-* or even just *-o-*, and the transitive forms *-ar-*, *-av-* (also *-er-*, *-ev-*), *-ker-*, *-arker-*, *-avker-*. A younger, alternative intransitive marker is *-áv-*, derived from the auxiliary *av-* 'to become', from *av-* 'to come'. Transitive affixes attach to intransitive, de-nominal and de-adjectival roots to form transitive verbs: *sikhlj-ar-* 'to teach' (*sikhlj-* 'to learn'), *dar-av-* 'to frighten' (*dar-* 'to

fear'), *bar-av-* 'to enlarge' (*bar-o* 'large'). In some dialects they mayalso attach to transitive verb roots to form causative verbs: *dar-av-av-* 'to cause somebody to frighten'.

Borrowed verb stems are integrated by attaching a derivational vowel which mediates between the stem and tense-aspect inflection. The original inventory of loan verb adaptation markers appears to have consisted of a combination of a Greek-derived affix representing the inflectional class of the Greek verb – this model was set, apparently, during contact with Greek in the Byzantine or 'Early Romani' period – and an inherited verb-derivational marker that indicated transitive or intransitive valency. Traces of this old system still appear in some dialects: *vorb-is-ar-el* 'he speaks' (*vorb-* from Romanian 'to speak', *-is-* representing the Greek aorist, *-ar-* transitive derivation), *beklet-is-ker-d-jom* 'I waited' (Turkish *bekled-*, Greek aorist *-is-*, inherited transitive *-ker-*, past-tense *-d-*). Most dialects have simplified the inherited system and retain just one or two loan verb adaptation markers, often representing either derivational affixes (*-ev-*, *-er-*) or Greek-derived inflectional markers (*-iz-*, *-in-*) or combinations of both (*-is-ar-*, *-is-áv-*). In a number of dialects there are tendencies to incorporate some loan verbs along with their original tense-aspect and person inflection. This is common especially with modal auxiliaries. With lexical verbs the system is consistent especially in the Balkan and Black Sea dialects of Romani that have been under considerable influence of Turkish (and other Turkic languages, such as Tatar). Here, Turkish verbs often carry their entire Turkish inflection, resulting in an etymological split in inflection classes.

As in other Indo-Iranian languages, Romani verbs show present and past stems. On the whole, the forms are similar to the present forms and only a few verbs show a distinction, e.g. *sov-* 'to sleep', past *su-t-*; *per-* 'to fall', past *pe-l-*. The characteristic feature of the past (also: perfective) stem is usually just the added augment or perfective marker, as in *ker-* 'to do', past stem *ker-d-*. The perfective marker derives from the historical marker **-t-* which has taken on the forms *-t-*, *-d-*, and *-l-*, depending on environment (i.e. the final segment of the verb root). The present stem shows a split between vocalic roots and consonantal roots, which affects the shape of the person or subject concord marker that immediately follows it: *xa-s* 'you eat', but *phen-es* 'you say'. In some dialects, new inflection classes are formed through contraction of the intransitive derivation (*dikh-el* 'he says', *dikhj-ol* 'it is seen') or the loan verb adaptation marker (*vorb-il* 'he speaks'). In the past stem, inflection classes are identified by the distribution of the individual perfective markers named above. The group of mono-consonantal verb roots constitutes an independent inflection class. Additional markers/classes are verbs of motion and change of state, including mediopassive, which often take the perfective marker *-il-*, and those that end in *-a* and express senses/emotions, such as *dara-* 'to fear' and *asa-* 'to laugh', which have a variety of perfective markers, including *-ndil-* (thus *asa-andil-jom* 'I laughed'). The present and past stems each have separate set of subject concord markers. Those used with the

present verb stem are direct descendants of Old and Middle Indo-Aryan person markers; Romani is one of the most archaic New Indo-Aryan languages in preserving this set. The set of person suffixes used with the past or perfective stem derive from enclitic oblique pronouns that originally identified the agent in an ergative construction and became re-analysed as subject concord markers in a new conjugation, a development that is typical of many of the Indo-Iranian frontier languages.

The verb has two external tense markers, that follow person endings: The suffix *-a* attaches only to verbs in the present stem and indicates either the indicative present-future, or specifically the future or conditional. The suffix *-as* (also *-ah, -ahi, -s*) indicates a 'remote' perspective; combined with the present stem it forms the imperfect or habitual, and combined with the perfective or past stem it forms the pluperfect or counterfactual. Most Romani dialects lack an infinitive, though in some dialects, especially in central Europe, there is a tendency to generalise just one finite form in modal constructions and so to drop agreement between identical subjects in the main and complement clause. Many Romani dialect preserve a present-participle (gerundial) form of the verb; otherwise, verbs clauses are generally finite. The finite form of the verb without any extension serves as a subjunctive.

Apart from the verb 'to go', the copula is the only verb in Romani that is genuinely suppletive. Future and subjunctive forms of the copula derive from the verbs 'to become', 'to come' or sometimes 'to stay'. The present set of the copula makes use of the same set of person concord markers as the perfective stem in lexical verbs, indicating in all likelihood that the present copula derives from forms that had once been used for the past tense, while now the past tense of the copula is forms exclusively by adding the remoteness marker *-as*, without forming any separate perfective stem.

4.   Syntax

4.1.   Word order type and alignment

Romani generally displays the properties of a SVO language with a flexible, pragmatically determined word order. This flexibility is reminiscent of Greek. In declarative clauses, there are several slots to the left of the verb which may be occupied by (multiple) topics. The subject appears to the left of the verb when it is topical and when it helps establish the perspective on the proposition. The subject appears to the right of the verb when there is a tight, consequential connection between the sentence and the preceding context. Relative clauses follow their head nouns and contain resumptive pronouns wherever there is reference to the head in a non-subject role. Pro-dropping occurs especially in the chaining of subjects, and in this respect Romani is a pro-drop language to a limited extent. Clauses are

generally finite and subordinations are introduced by conjunctions, which are typically derived from interrogatives or local relations expressions, or borrowed. Like other Balkan languages, Romani shows a split between the factual and non-factual complementisers. The difference is indicated by the choice of conjunction that introduced the complement clause, as well as in the choice of verb mood – subjunctive in non-factual clauses, indicative in factual clauses. The subject tends to precede the verb in factual clauses, but follows the verb in non-factual clauses.

Romani displays accusative alignment. However, its verbal system shows some traces of ergativity in earlier stages of the language. Thus the past-tense conjugation is derived from a set of oblique enclitic pronouns which once indicated the agent. IN the more conservative dialects primarily in southeastern Europe, there is still a tendency for intransitive verbs of motion and change of state (including mediopassives) to show participial (adjectival) agreement in the third person, rather than person inflection (deriving from agentive morphology), thus: *ov/oj kerd-jas* 'he/she did', but *ov gel-o* 'he went', *oj gel-i* 'she went'.

4.2.   Order inside the nominal phrase

Attributes generally precede their head nouns. Definite and indefinite articles and demonstratives usually occupy the first slot in the noun phrases, followed by numerals, then by genitive-possessive nouns, then by adjectives. For emphasis, adjectives and genitive-possessive nouns may be placed after the noun, and in some dialects this is true of demonstratives, too.

Table 1: Typological features of Romani

Structural features	realisation in Romani
Animacy	differential object marking
Gender	two – masculine vs feminine – in nouns, singular attributes, and non-finite verbs in the 3rd person singular
Definiteness	pro-clitic definite article and indefiniteness marker
Number	singular, plural
morphological case	present
syntactic case	marked by adpositions
possessive constructions	1) attributive 2) predicative: no etymological *habeo*-verb 3) external possession
Aspect	non-perfective, perfective
Tense	Future, remoteness
Copula	suppletive in subjunctive and future

Mood	morphological indicative, subjunctive and imperative;
Passive	morphological: intransitive derivation (mediopassive), reflexive/passive; analytical.
Adpositions	prepositions
Adverbs	small number of primary adverbs
Adjectives	agree with their heads (in attribution and predicatively) in gender, case and number; analytical and synthetic comparatives coexist
word order	pragmatically free SVO with Adj N, Gen N, Prep N, Dem N, Num N and N Rel
Alignment	accusative

5.   Dialects

Geographical dispersion and the influence of diverse contact languages have led to the emergence of considerable difference among the dialects of Romani. Most of these differences are thought to have emerged during the period of settlement in the various regions of Europe, between the migrations of the fifteenth century and the period of dense attestation in the eighteenth century, by which time the overall picture of dialect differentiation is very similar to the present-day pattern. Support for this hypothesis can be found in the geographical distribution patterns of isoglosses. A dense cluster of isoglosses, referred to as the Great Divide (Matras 2005), separates Romani dialects on both sides of what was, during the relevant period, the border zone (and war zone) between the Austrian and Ottoman empires. There are both innovations and conservative features on both sides of the Divide. Other geographical distribution patterns separate a 'Black Sea' zone of innovations in and around Romania and northern Bulgaria, a western zone with its centre in an around Germany, as well as a centre-periphery dynamic. On the basis of the developments mainly in morphological paradigms it is possible to apply a classification grid to Romani dialects which recognises the following groups: Northwestern, Northeastern, Central, Vlax (Transylvanian-Vojvodina), and (southern) Balkan (cf. Elšík & Matras 2006, Boretzky & Igla 2004). Typical isoglosses that divide the Romani-speaking landscape pertain to the status of prothetic segments (*j-ov*, *v-ov* vs. *ov* 'he', *a-sa* vs. *sa-* 'to laugh'), palatalisation and affrication (*tikno* vs. *cikno* 'small', *gi* vs. *dži* 'soul', *geljas* vs. *gejas* 'gone'), the distribution of a set of lexical variables, and reduction and analogical formations within morphological paradigms. Once speakers adjust to a monolingual mode of discourse, it is generally possible for Romani speakers east of the Great Divide (from Greece to the Baltics) to understand one another, while the (much smaller) population of speakers to the west of the Great Divide speak dialects that are more fragmented.

References

Boretzky, Norbert and Birgit Igla.   2004. *Kommentierter Dialektatlas des Romani.* Wiesbaden: Harrassowitz.

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

Matras, Yaron 2002   *Romani: A linguistic introduction.* Cambridge: Cambridge University Press.

Matras, Yaron.   2005. The classification of Romani dialects: A geographic-historical perspective. In: Barbara Schrammel, Gerd Ambrosch, Dieter W. Halwachs (eds.), *General and applied Romani linguistics*, 7-26. Munich: Lincom Europa.