

Problems in the Codification of a Standard Romani Literary Language

Victor A. Friedman

University of North Carolina, Chapel Hill

from: *Papers from the Fourth and Fifth Annual Meetings: Gypsy Lore Society, North American Chapter.*
 New York: Gypsy Lore Society, 1985. 56-75.

In this article, I shall discuss some concrete linguistic problems connected with the formation and standardization of a Romani literary language as reflected in the Romani grammar entitled *Romani Gramatika* (Macedonian *Romska Gramatika*, henceforth referred to as RG) by Saip Jusuf and Krume Kepeski published in Skopje 1980. The book contains parallel texts in Romani and Macedonian and is undoubtedly the most ambitious attempt of its kind to be published so far. The express purpose of RG is the creation of a Literary Romani for use by Roms in Macedonia, Kosovo, and adjacent parts of Serbia, with a view to the creation of Romani-language schools in these areas and to the use of this literary standard as a basis for the creation of a Romani literary language for use by Roms in general. The language of the grammar is based on the Arlija dialect of Skopje, although Jusuf makes frequent use of his native Dzambaz dialect, especially when citing Romani forms in the Macedonian text, and occasionally Gurbet and Burgudži forms are also mentioned.

In its intent and execution, RG can be compared to the pioneering works of the Macedonian patriot Górgi Pulevski, especially his polyglot dictionaries and his grammar entitled *Slognica recovska* published at the end of the nineteenth century (Ristovski 1974:33-181). Pulevski wrote at a time when Macedonians were acquiring a national consciousness as a distinct ethnic and national group. His dictionaries were in fact conversation manuals in which he discussed such topics as language and national identity in parallel Slavic, Turkish, Albanian, and, in the first edition, Greek texts, i.e., he used the language of his intended audience (Macedonian), the official language of the state (Turkish, as Macedonia was still a part of the Ottoman Empire), the language of his audience's most numerous neighbors within the state (Albanian), and, in the first edition, the language of his audience's church (Greek). Pulevski's grammar was intended as a first step toward the establishment of a Macedonian literary language.

While the comparison between Pulevski's work and RG must be made with great care, as many obvious differences do exist, there are similarities that are striking and also instructive in setting RG in an appropriate historical and sociolinguistic context. Like Pulevski, Saip Jusuf is not a professional linguist by training, but unlike Pulevski he co-authored RG with the author of other grammars, viz., Krume Kepeski, who wrote the first standard high-school grammar of Macedonian, a revised edition of which is still in use (Kepeski 1972).¹ As in Pulevski's work, so in RG we have the use of the language of the intended audience and that of their most numerous neighbors (in most regions), which is also an official language of the state, i.e., of the Socialist Republic of Macedonia. As in Pulevski's work, so in RG, the authors are attempting to treat a language with numerous dialects and sub-dialects that have interacted with one another and for which no standard has been officially established at the time of publication.² In both works, compromise was attempted by means of varying degrees of free variation in forms and usage.

Unlike Pulevski, however, RG did not have to contend with the type of archaizing, ecclesiastical tradition that existed for Slavic, although the policy of RG, supported by the World Romani Congress (Puxon 1978/9:69), concerning the borrowing of words from Hindi can be seen as a related puristic phenomenon comparable to, e.g., the Turkish puristic movement of the Atatürk reforms in which words of Arabic and Persian origin were replaced by words borrowed from or coined on the basis of related Turkic languages. Nevertheless, historical circumstances preclude the archaizing/popularizing diglossia which, e.g., threatened Bulgarian and still plagues Greek. The problem of combining literary dialects (in Romani, nascent literary dialects) used by different populations, or rather of choosing one type as basic while incorporating elements from others -- a process officially completed for Albanian only in 1972 -- will be influenced by more factors than can be considered here, but many of the basic types of problems facing the codification of Literary Romani are comparable to those of other languages.

As was indicated earlier, the intended audience of RG is the Romani population of Macedonia and adjacent areas. It is not, therefore, intended for use by non-Roms to learn Romani but rather to teach Roms to perceive their native language as a means and object of formal education. Nonetheless, a high degree of descriptive adequacy is a major desideratum even for such a grammar. This is especially true of a grammar attempting to promote the establishment of a literary standard. The remainder of this article will be devoted to an exposition of specific linguistic problems facing the creators of a Romani literary standard based on the data in RG.

1.0 *Orthography and Phonology.* With certain exceptions to be noted below, RG employs a Cyrillic and

¹ We should note here that Macedonian was not officially recognized as a literary language until 1944.

² The question of whether Romani is to be considered a single language with numerous dialects or as a group of closely related languages is not of immediate concern to this article. The most commonly held opinion among linguists is that Romani is to be treated as a single language (cf., Kochanowski 1963:184-92; Hancock 1975:26; Ventcel' and Cerenkov 1976:283), despite varying degrees of mutual intelligibility. Questions of the definition of language and dialect involve more factors than mere intelligibility, as can be seen from such classic examples as the mutually unintelligible *dialects* of Chinese or the mutually intelligible *languages* of Scandinavia, but the discussion of the relationship among language, dialect, ethnic identity, and national identity are beyond the scope of this article. For our purposes, we shall adopt the commonly held view of linguists just alluded to.

a Latin orthography based on the Cyrillic orthography of Macedonian and its standard Latin transcription as practiced in the Socialist Federated Republic of Yugoslavia. While this is slightly at variance with the orthography proposed by Hancock (1975:41-42) and that of the World Romani Congress (Kenrick 1981), it is a practical first step for the Roms of the Socialist Republic of Macedonia and Yugoslavia in general, and it is reasonably consistent with the pronunciation of the Romani dialects that it represents. On the basis of the data in RG, we can identify the following specific problem areas in the standardization of a literary Romani sound system and its representation:³ the position of schwa in the system and its representation; differentiation of the voiceless velar and uvular fricatives and the glottal approximant (glide); use and spelling of jot; phenomena connected with palatal stops; velarized vs. nonvelarized lateral approximants; voicing phenomena; word division; stress.

1.1 *Schwa*. RG uses the orthographic representation *ä* for this sound in both its Latin and Cyrillic orthographies. As the representation of this sound tends to be peculiar to the orthographies of individual languages using a separate letter to represent it, e.g., Romanian *a*, Albanian *ë*, Bulgarian *ѐ*, Macedonian *ä*, and as the international phonetic symbol *ɨ* does not occur on most typewriters, the choice of yet another separate letter for Romani is not without justification. The position of schwa in the sound system of Romani, particularly the dialects of RG, however, is another question. All dialects of Romani possess the five vowels *a, e, i, o, u*, but many dialects, including some dialects of Balkan Romani, have only those five vowels (Ventcel' and Cerenkov 1976:296). Even in the dialects of RG, schwa appears to have a marginal position at best. Aside from its extreme rarity and apparent limitation to non-initial and non-final position, it is used inconsistently in certain words of both native and foreign origin. Thus, for example, while *akäl* "intelligence" (Turkish *akıl*) is consistently spelled, *sasträn* "iron" (Sanskrit *sāstri* [noun], *sāstrin* [adjective]) and *garäpi* "poverty" (Turkish *garip* from Arabic *garib*) also occur as *sastri*, *sastrun*, *garipi*, in RG and the dialects on which it is based. The problem of schwa in Romani dialects and the literary standard is thus clearly in need of further elucidation.

1.2 *Voiceless dorsal non-plosives*. In the various dialects of Romani there are three types of voiceless sounds made at the back of the mouth and throat and not involving full occlusion: the velar fricative (as in Slavic); the uvular fricative (as in Arabic, Hebrew, Georgian, etc.); and the glottal approximant (English *h*) (cf. Ventcel' and Cerenkov 1976:294). RG recognizes two of these: the velar (Latin and Cyrillic *x*) and the glottal (Latin *h*, Cyrillic *x'*). This orthography is somewhat at variance with the standard Yugoslav Latin transcription of Cyrillic, where *h* represents the velar fricative, but it is consistent with the value of *h* in many other orthographies and phonetic transcriptions, and the value of *x* in the Cyrillic orthography is consistent with its value in the Slavic languages employing it. In the colloquial Romani of Macedonia, one can hear all three types of dorsal non-plosives in the pronunciation of the same word by the same speaker at different times, indicating that these sounds are in free variation. Another indication that the difference between *h* and *x* may not be distinctive is the fact that the same word or root is spelled with both *h* and *x* in RG, e.g., *xiv*, *hiv* "hole," *xor* "depth" but *hordaripe* "deepening," *xramondo* "letter" but *hramondikano* "written," etc. This problem is also alluded to by Kenrick (1981:3), who suggests *h* for *x* and uses *hav/xav* "I eat" as his example, although RG consistently uses *hav* in its Latin orthography while the Cyrillic varies between *хав* and *х'ав*. It can thus be seen that the distinctiveness of these sounds and their occurrence in various words has yet to be established.

1.3 *Jot*. Problems with jot involve three types of phenomena: the loss of intervocalic *s* in the Arlija dialect; the spelling and pronunciation of intervocalic jot in various grammatical and derivational forms in all dialects; and jot vs. *i* in final and preconsonantal position after a vowel.

The change of *s* to *h* in general or in certain restricted grammatical environments is well known in a wide range of Romani dialects (Ventcel' and Cerenkov 1976:285). In the Arlija dialect of Yugoslavia, *s* → *h* → *j/∅* intervocalically in the instrumental singular, second person singular and first person plural present tense endings, all of which have the shape *-sa* in other dialects, e.g., *ac devle[j]ja* (< *devlesa*) "good-bye" (literally "stay with God"), *tu/amen dza[j]ja* (< *dzasa*) "you/we go," etc. The behavior of this jot is the same as that of other intervocalic jots in Arlija and other Romani dialects of Macedonia, and so problems connected with its presence or absence will be treated in the following paragraph. We shall conclude this paragraph by pointing out that RG implicitly favors Yugoslav Arlija over Dzambaz (which preserves *s*, as does the Arlija dialect of Bulgaria) in both its exposition and usage. Although RG mentions the instrumental desinence *-sa* it does not mention the 2sg./1pl. *-sa* in the relevant grammatical expositions, and it usually employs Arlija forms in the actual text.

The question of the pronunciation and spelling of intervocalic jot is a vexed one for a number of languages, notably Macedonian and Albanian, which are particularly relevant not only because they are both familiar to the intended audience of RG but also because the official codification of their current orthographies has taken place since World War Two. Thus, for example, in Macedonian *j* is never written intervocalically before *i* and *e* (unless *e* is preceded by two vowels, e.g., *znae* "he knows" but *faoje* "foyer") but it is always written between *i* and *a* (i.e., *ija*). Its orthography in other intervocalic environments can depend on morphological conditioning, e.g., *laat* "they bark" but *laja* "it barked," or *vior* "whirlwind" but *Marijo* "O Maria."

³ It should be stressed here that the problem areas discussed throughout this article are limited to a standardization based on the dialects of RG. Problems arising from the incorporation of features from other dialects are not germane to our discussion, although some of them will undoubtedly affect the codification of Literary Romani. Even a standard based on dialects other than those of RG would have to take the problems raised by it into account.

This all reflects orthographic convention rather than actual pronunciation (v. Tosev et al. 1970:10-12). Similarly, in Albanian specific rules govern the orthography of intervocalic jot, e.g., *e tija* "his" (from *tij*) but *e mia* "mine" (from *mi*), *anija* "the vessel" (from *anije* "vessel") but *shtëpia* "the house" (from *shtëpi* "house"), etc. (v. Kostallari et al. 1974:74-79); this despite identical pronunciations. In RG, the problem of establishing conventions for the orthography of intervocalic jot has not been solved except in the sequence *ija*, e.g., *dzamija* "mosque," which consistently follows Macedonian practice. Elsewhere, forms with and without jot are used in different occurrences of the same word, e.g., *arlijengoro/arlijengoro* "Arlija," *dzamie/dzamije* "O mosque," cf., also, e.g., *katilukoa* "with cruelty" but *cakalukoja* "by cleverness," etc. As in Macedonian and Albanian, so in Romani, the degree of distinct pronunciation of intervocalic jot will vary with environment, speaker, and occasion, but the fixing of an orthographic convention for its presence or absence in writing will be one of the tasks of literary codification.

The spelling and pronunciation of jot vs. *i* in post-vocalic final and pre-consonantal position can be similarly problematic. In Macedonian and Albanian, the different spellings distinguish grammatical forms which in ordinary conversation would sound virtually the same, e.g., Macedonian *slucaj* "event"/*slucaj* "events", Albanian *harroj* "I forget"/*harroi* "he forgot." RG is consistent in writing final jot in some words, e.g., *thaj* "and" but inconsistent in others, e.g., *saj/sai* "it is possible," *muj/mui* "mouth." Such alternative spellings are common in many works in and on Romani.⁴ Given the grammatical structure of the dialects of RG, it might be possible to establish an orthographic convention distinguishing masculine nouns such as *sej* "thing," *puj* "chicken," *nilaj* "summer" from feminine nouns such as *lej* "river," *kuj* "elbow," *phabaj* "apple" by writing the latter with a final *-i*, which is a native Romani feminine marker. This, however is a problem for future codification.

1.4 *Palatal stops.* These sounds are represented in RG by the Macedonian Cyrillic *ќ, ѓ* and the Latin digraphs *kj, gj*. According to RG, the Romani sounds are closer to the Macedonian sounds represented by the Cyrillic letters *ќ, ѓ* than they are to the Serbo-Croatian sounds represented by the Cyrillic *ћ, њ* and the Latin *c, d*. It is certainly the case that in the pronunciation of palatal or palatalized stops or affricates the various dialects of Romani often agree most closely with the pronunciation of the non-Romani languages or dialects with which they are in closest contact (cf., Ventcel' and Cerenkov 1976:293, Kochanowski 1963:99). The dialects of RG are typical in this respect (cf., Kochanowski 1963:84 on dialectal variation comparable to that of non-Romani contact dialects.) As a result of the similarity of the dialects of RG to Macedonian dialects (and also, one might add, to some Turkish and Albanian dialects of Macedonia) in this respect, two questions arise with regard to the orthographic representation of palatal stops: firstly, the fronting of *k, g* before front vowels (*i, e*), and secondly the palatal mutation of *t, d*, followed by jot.

In the dialects of RG, as in Macedonian and West Rumelian Turkish, velar stops are automatically fronted to palatal stops before front vowels and jot. Thus, for example, in RG we have *ko gav* "to the village" but *kji diz* "to the town." Since this fronting is automatic, one could also write *ki diz* without changing the pronunciation. In fact, RG uses both *kj, gj* and *k, g* indiscriminately before front vowels, e.g., in the preposition just mentioned, in the dative suffix (sg *ke/kje* pl *ge/gje*), in various lexical items, e.g., *kerdo/kjerdo* "done," etc. In Macedonian, orthographic rules prescribe the writing of *k* before *i* and *k'* before *e* (except in certain words and proper names, v. Tosev et al. 1970:19), despite the fact that the pronunciation is always that of a dorso-palatal stop. Some rule will also have to be formulated in the codification of Literary Romani.

In the dialects of RG, as in Macedonian, stem-final *t, d* automatically change to the corresponding dorso-palatal stops when followed by jot (which may come from a non-syllabic *i*), e.g., *buti/bukjaa* ↔ *bukjasa* "work nom. sg./instr. sg.," *kerdo/kerdjum* "done/I did," etc. (cf., Macedonian *cvet* "flower," *cvekâr* "flower-seller," etc.) In other Romani dialects, this palatal mutation does not occur; rather *t, d* are simply jotted or palatalized or, depending on the shape of the ending in the given dialect, left plain, e.g., Russian Romani *kerd'om*, Lovari *kerdem* "Ibid.;" cf., also the form *vakerdjas* 'it said' (Kenrick 1981:3), which in RG would be *vakergja* or *vakjergja* (from *vakerdo* "said"). Here the usage of RG is consistent with both dialectal pronunciation and Macedonian orthographic practice, but the rules are not stated anywhere in RG and the resulting orthographic representations could prove confusing to Roms elsewhere.

1.5 *Laterals.* According to RG, the digraph *lj* is pronounced like the Macedonian *l*, and the sound represented by the letter *l* is soft (alveolar) before front vowels and hard (velarized) elsewhere. In Macedonian, *l* (Cyrillic *л*) is pronounced according to this description, and *lj* (Cyrillic *љ*) is restricted to indicating clear (alveolar) [l] in environments where the letter *l* (л) would be pronounced as a velarized lateral, e.g., *ljubov* [lúbof] "love," *biljbilj* [bílbil] "nightingale."⁴ The Macedonian *lj* is never written before *i* or *e*. In RG, we have both *lj* and *l* before front vowels in alternate spellings of the same words, e.g., *ljil/lil* "book." If the phonetic situation is the same as in Macedonian, then Literary Romani could justifiably use the same orthographic conventions. If, however, there is a phonological contrast between the sounds represented by the letters *lj* and *l* before front vowels, then a consistent distinction should be made.

1.6 *Voicing phenomena.* The dialects of RG have the same type of voicing phenomena as in Macedonian, viz., final distinctively voiced consonants are devoiced and all consonants entering into the voiced/voiceless opposition, when occurring in a consonant cluster, will be either all voiced or all voiceless depending on the last consonant, e.g., Romani *dad* [dat] "father"/*dadestar* [dadéstar] "abl. sg.," *pandz* [panc]

⁴ This is in contrast to Standard Serbo-Croatian, where *l* always represents a clear alveolar lateral approximant and *lj* represents a palatal one. This is also in contrast to the speech of some younger Macedonians, where *l* is velarized in all environments.

"five"/*pandzto* [pancto] "fifth"/*pandzende* [panʒende] "five, loc. sg.," etc.⁵ The orthography in RG is generally consistent in portraying underlying voiced and voiceless consonants in environments of neutralization, as indicated in the examples just given, although no explicit rules are stated, and occasional slips do occur, e.g., the spelling of *dat* for *dad*. The spelling of underlying consonants in environments of neutralization is especially desirable for Literary Romani not only because it gives greater predictability but also because there are dialects where, e.g., final devoicing does not occur (Kochanowski 1963:58).

1.7 *Word division*. RG gives rules for writing certain particles and clitics as separate words or as prefixes, but a number of conventions remain to be established; see section 2.3 for an example.

1.8 *Stress*. It has been observed (e.g., Ventcel' and Cerenkov 1076:297) that the stress patterns of a number of Romani dialects tend to conform to those of the chief non-Romani contact language. Thus, for example, some Romani dialects of Germany, Hungary, and Czechoslovakia tend to have initial stress, those of Poland tend to have penultimate stress, etc. The Balkan Romani dialects, however, including those of RG, have tended to conserve what must have been the original oxytonic stress pattern. In its section on stress, RG states that Romani stress is free although it tends to be final, and that it is mobile within a paradigm. However, the examples adduced -- *Róm*, *Romá*, *roméskoro*, "Rom nom. sg., nom. pl., gen. sg.," *chingjórav*, *chingjóraja* "get tired 1 sg., 1 pl." -- while demonstrating that Romani stress cannot be defined as fixed final, penultimate, or antepenultimate (as opposed to, e.g., Literary Macedonian, where stress is defined as fixed antepenultimate except in certain recent loanwords), do not demonstrate mobility within a paradigm. In both instances, as is true of native Romani words in general, the stress is fixed on the stem-marker (thematic vowel). Although RG does not formulate any rules for predicting stress, the dialects on which it is based do in fact display a high degree of predictability in stress assignment. In general, stress will be final except in the following:

1. Present tense stress is on the stem vowel, e.g., *peráva/peráv* "I fall," *perél* "he falls" (see section 2.4).
2. Compounds with *ováv* "become" do not stress the compounding verb in the past tense (also the present in some dialects), e.g., *bokhálilo* "he got hungry" (*ilo* is the compounding past tense form).
3. The oblique forms of all nouns stress the stem vowel, i.e., the syllable before the case marker, e.g., *Roméške* "Rom dat. sg.," *Roméskoro* "Rom gen. sg.," *lafíste* "word dat. sg.," *lafískoro* "word gen. sg."
4. Recently borrowed substantives ending in a vowel stress the syllable before the vowel in non-oblique forms, e.g., *láfi* "word," *lafíja* "words."
5. Certain prepositions, conjunctions, etc., have nonfinal stress, e.g., *hémako* "although."

Assignment of stress is one of those areas in which many literary languages, e.g., Bulgarian, English, and Russian, permit a degree of variation within the limits of certain norms. This will undoubtedly also be the case for Literary Romani, although the structure of the languages is such that some basic patterns can be defined. While the indication of stress is desirable in reference and pedagogical tools, a codifying grammar would only need to formulate the basic rules for predicting stress assignment and the major classes of exceptions.

2.0 *Morphology*. The section on morphology is the most extensive, occupying more than half of RG. RG goes to considerable lengths to present traditional definitions of traditional grammatical terms in keeping with the goal of providing Roms with the means to think of their language as a legitimate object of formal education. The main problems facing the codification of Literary Romani are connected with the standardization of paradigmatic forms and usage. In this respect, RG has illustrated the wide range of choices by using various forms of the same words and endings in both its text and its grammatical exposition. A key issue here is the matter of dialect choice, but some of the variation is not dialect specific. In the following sections, we cannot hope to cover every single problem which will ultimately engage the attention of the codifiers of Literary Romani, and so we shall focus on some of the most important problems raised by the data in RG.

2.1 *Substantives*. The following phenomena are among those in Romani substantival morphology that will require clarification in the future codification of Literary Romani: recent masculine borrowings with nontruncating stem vowels; jotating vs. non-jotating feminine stems; contracting or truncating masculine and feminine stems; the forms of the accusative, genitive, instrumental and vocative cases; derivation and inflection of abstract nouns in *-pe* and derivation in general.

In the dialects of RG there are two types of masculine substantives ending in a vowel: older, truncating stems and newer, nontruncating stems. Although the latter type of substantive is amply illustrated in the text of RG, the paradigms in the exposition of declension exemplify only the first type, e.g., *vogi* "stomach," *vogjestar* "abl. pl.," *vogja* "nom. pl.," *vogjendar* "abl. pl." The second type could have been exemplified by a word such as *lafi* "word," *lafistar* "abl. sg.," *lafíja* "nom. pl.," *lafijendar* "abl. pl." In both the teaching of Romani in schools and in its literary codification it will be necessary to give an explicitly complete account of its grammatical structure. The adaptation of new words is one of those areas of the Romani grammatical system that varies significantly from dialect to dialect and is simultaneously one whose codification will be of major concern to the establishment of Literary Romani.

A similar problem of nonpredictability of the oblique stem on the basis of the nominative singular is to be found in feminine substantives. Of those substantives ending in *-i*, some lose the *i* in the oblique cases and some preserve it in the form of jotation, e.g., *píri* "pitcher," *pirate* "loc. sg.," *pirende* "loc. pl.," but *Romni*

⁵ We should note here that /v/ is affected by the rules of final devoicing and regressive assimilation of devoicing but does not itself trigger regressive assimilation of voicing in Macedonian. In the dialects of RG, /v/ behaves like a sonorant with regard to voicing rules and can also be realized as [w], especially in word final position.

"Rom, woman, wife," *Romnjate* "loc. sg.," *Romnjende* "loc. pl." Similarly, feminine substantives ending in a consonant may have jotating or nonjotating stems, e.g., *balval* "wind," *balvalate*, *balvalende* but *diz* "town," *dizjate*, *dizjende*, and some stems can be either jotating or nonjotating, e.g. *chib* "language, tongue," *chibate/chibjate*, *chibende/chibjende*. In its text, RG has examples of all these types of substantives, including both forms of stemš such as *chib*, which admit variation. In its exposition of declension, however, it only gives tables for nonjotating stems and does not discuss jotation. As was mentioned in the paragraph above, explicit completeness, as well as the determination of membership in a given paradigmatic class, will be a major goal of literary codification.

A similar problem is that of feminine and masculine substantives ending in a vowel plus jot, e.g., *muj/mui* "mouth" (cf., section 1.3). The substantives may or may not undergo contraction and vocalic alternation as illustrated by the text but not the exposition of RG, e.g., *muj* "mouth," *moste* "loc. sg.," *daj* "mother," *dajakiri* or *dakiri* "gen. sg." Our comments in the two preceding paragraphs also apply here.

The use and forms of the various cases represent significant problems for the codification of Literary Romani. We have already discussed the problems connected with the choice of forms of the instrumental singular (-*sa* vs. [-*j*]a) in section 1.3, q.v. Although RG only mentions the vocative singular in -i[*j*]e for feminine consonant stems in its exposition of declension, e.g., *balvalie* "O wind," in its exposition of the definite article it also cites the ending -*orije/-orija*, e.g., *darorije/darorija* "O fear," although this appears to involve the (derivational) diminutive suffix -*ori* (masc. -*oro*). RG also cites the tendency among some Roms to use the nominative for the vocative.

In its exposition of the genitive, RG mentions only the long masculine form, e.g., *Romeskoro* "Rom's," *Romengoro* "Roms'." There are two additional problems facing the codification and analysis of Literary Romani not mentioned in RG, viz., the short genitive and the phenomenon of agreement with the head noun due to which many analyses of Romani do not consider this a case form but rather a derived adjective (cf., Ventcel' and Cerenkov 1976:305).

Elsewhere, Jusuf (1981:9-10) has described the dialectal differentiation in the Balkans (and, we might note, elsewhere) between the disyllabic genitive ending -*koro*, -*kiri*, -*goro*, etc., and the monosyllabic -*ko*, -*ki*, -*go*, etc., the former being particularly characteristic of the Arlija dialect. In the text of RG, both types of genitive are used, e.g., *o arliengo/arliengoro vakjeribe* "the Arlija dialect" (literally "the Arlijas' dialect"), but in the exposition of declension only the former is cited.

The treatment of what we have been calling the *genitive* as a case form, following RG, is particularly problematical. Unlike the other cases, the genitive agrees in gender, number, and case with its head, just as any adjective does, e.g., *o arliengoro vakjeribe* "the Arlija dialect," *e Romengjiri istorija* "Romani history/the history of the Roms," *e manusengjere chive* "the people's children," *e Romengjere chibjaa* "with the Romani language/the Roms' language." RG does not mention the phenomenon of agreement in its exposition of declension nor does it enter into the problem of the classification of these forms. Such questions will have to be treated in the codification of Literary Romani (cf., Grumet's paper in this volume).⁶

In the declensional tables of RG, all substantives are cited as having distinct accusatives except the singulars of masculines in -*pe*, e.g., *suzipe* "beauty," and feminines in -*ija*, e.g., *dzamija* "mosque." In actual usage in the text, however, the dialectš serving as the basis of RG appear to observe the animate/inanimate distinction described by, e.g., Ventcel' and Cerenkov (1976:301; cf., Kostov 1962:134-35; Sampson 1968:133), i.e., animate nouns have a distinct accusative while inanimates do not distinguish accusative from nominative, e.g., *dikhljum e basne* "I saw the rooster" (nom. *basno*) but *haljan mo maro* "you ate my bread" (nom. *maro*). This distinction, however, is not consistently observed in all the dialects forming the basis of RG, e.g., *dikhlja argatijen/argatija* "he saw [some] field hands" (acc./nom.)

In our discussion of problems connected with case and the standardization of Literary Romani, we can mention that the rules for case usage show considerable variation, one small example of which we have just cited in the paragraph above. Kostov (1962:135-39, 143-45) gives an impressive array of phenomena such as the use of the accusative as the pure objective case (the dative being reserved for concrete spatial and related functions, cf., Sampson 1968:133), variations in case government with different prepositions, etc. The text of RG contains relevant material for the examination of such phenomena, but it does not expound upon them in its sections on case and syntax. The description, analysis, and codification of case form usage will be a major task in the codification of Literary Romani.

The suffix -*pe* (-*be*, -*pa*, -*ba*; Sanskrit -*tvanna*, Prakrit -*ppaṇa*) is one of the most important and productive for forming abstract nouns in Balkan Romani and, *mutatis mutandis*, in other dialects. In some dialects, this suffix preserves the original -*n* in the oblique or in all case forms (v. Sampson 1968:75), but this does not appear to be the case in the dialects of RG (except for the vocative plural *suzipenalen* "O beauties"). In the text of RG, this suffix has the four possible forms just cited, e.g., *vakeribe* "dialect, speech," *khelipe* "dancing," *sikljoviba* "science, scholarship," *vazdipa* "raising, stress." In the declensional table of *suzipe* in the exposition of declension, RG attempts to unify the vocalic variations by giving the oblique štem vowel

⁶ The phenomenon of the agreement of a genitive with its head is known in other languages, e.g., the so-called double declension of Old Georgian, in which the genitive is added onto the case ending of its head when following that head, e.g., *mep-e* "king," *kal-i* "daughter" (-*e* and -*i* are nominative case markers; in actual orthography all case markers are written as parts of the words to which they are attached), -*sa* "dative case," -*isa* "genitive case" hence *mepisa kalsa* "to the king's daughter" but *kalsa mepisasa* "to the daughter of the king" (Vogt 1971:36).

alternatively as *-e-* in some cases (e.g., dat. *suzipeskje*) and as *-a-* in others (e.g., loc. *suzipaste*). This does not represent a systematic alternation, however, and in the codification of Literary Romani it will be desirable to settle upon a single paradigm.

A brief comment can be made here on the treatment of derivational morphology in general. There are two basic methods of organizing an analysis of derivation: according to the original part of speech to which the base belonged and according to the form and function of the suffix.⁷ Organization by part of speech such as we find in RG is the clearest method of demonstrating that derivation can produce different parts of speech from a given lexical root or stem. To show how the process of derivation modifies the meaning of a lexical root or stem in a transparent and predictable manner, however, organization by suffix is more useful. Thus, for example, in RG *khelipe* "dancing" (*khelēl* "he dances"), *manusipe* "humanity" (*manus* "person"), *cakalipe* "cleverness" (*cakali* "clever") and *dujipe* "duplication" (*duj* "two") each occurs in a different group of derivatives according to the part of speech from which it is derived, whereas the ultimate codification of Literary Romani will find it more useful to list the productive derivational suffixes and then subclassify the types of words to which they apply by part of speech.

2.2 Adjectives. There are four items of particular importance in the codification of the Literary Romani adjectival system: comparison; treatment of uninflected adjectives; derivation of new adjectives; and choice of possessive pronominal adjectives.

The treatment of comparison of adjectives in RG is quite thorough, especially in the context of the audience for which it is intended. The codified literary language will need to choose norms, but the text of RG attempts to give equal time to various forms. The Gurbet dialect has preserved the synthetic comparative in *-eder*, e.g., *but* "much," *buteder* "more," while the other dialects of RG use the analytic comparative in *po-* borrowed from Greek or Slavic, e.g., *pobut* "more." The superlative is formed analytically with the prefix *em-* in Arlija and *maj-* in Dzambaz (and also Gurbet), e.g., *embut/majbut* "most." We can add to this two notes of correction and observation. Firstly, the prefix *em-* is derived from the Turkish superlative marker *en* "most," which is used in this same function, not from *hem* "and," which is also preserved in this form and meaning in Romani. Secondly, the prefix *maj-* is borrowed from the Romanian *mai* in this same function, so its use is not older than that of the Turkish word. Finally, in Macedonia, some Romani speakers also make use of the Slavic superlative marker *naj-*, e.g., *najbut*.

A distinction is made in Romani between agreeing adjectives in *-o* (masc. nom. sg.), *-i* (fem. nom. sg.) and *-e* (elsewhere) and nonagreeing adjectives in a consonant or other vowel, e.g., *o suzo Rom* "the handsome Rom," *i suzi Romni* "the beautiful Rom (woman)," *e suze Romestar* "from the handsome Rom," *e suze Roma* "the handsome Roms," but *i sukar Romni* "the good Rom (woman)," *o akālija Rom* "the intelligent Rom," etc. The distinction between agreeing and nonagreeing adjectives is not made explicit in RG, but as it exists in all Romani dialects (Ventcel' and Cerenkov 1976:308), it will be a relevant point for a complete literary grammar.

The derivation of new adjectives, like the derivation of new words in general, will be a major consideration in the codification of Literary Romani. This problem is not explicitly discussed in RG, but a number of possibilities are illustrated by its text, especially in the formation of language names and terminology pertaining to linguistics. On the basis of the data in RG, we can distinguish three broad types: unassimilated borrowings from Slavic, assimilated borrowings from Slavic, and genitives. Unassimilated borrowings are adjectives borrowed from Slavic which do not agree with the nouns they modify. Those in *-o* are based on the neuter singular form, e.g., *makedonsko chib* "Macedonian language," *literaturno chib* "literary language," *istorisko gramatika* "historical grammar," *rusko čhibo* "Russian language," cf., *romani chib* "Romani language," *romani gramatika* "Romani grammar." Nonagreeing adjective in *-lija*, e.g., *akālija* "intelligent," are borrowed from Turkish (*akilli*) but via the Slavic form of the derivational suffix (*-lija*). Words in the preceding class can also be assimilated as inflecting adjectives by means of suffixal *-n-*, e.g., *makedonskoni chib* "Macedonian language," *nacionalnone chibjake* "national language, dat. sg.," etc. Names of languages, etc., can take the form of genitives, both singular and plural, short and long, with truncating and nontruncating stem vowel, e.g.,

<i>istoriakiri gramatika</i>	historical grammar
<i>makedoniakiri chib</i>	Macedonian language
<i>rusoski chib</i>	Russian language
<i>dzambazəngoro vakeribe</i>	Dzambaz dialect
<i>germanijengiri chib</i>	Geřman language
<i>arlijengo vakeribe</i>	Arlija dialect
<i>araponge cifraja</i>	with Arabic numerals

The treatment of possessive pronominal adjectives in RG is rather fragmentary. The exposition mentions only two of the three types of variants for "my" and "your (sg.," used in the relevant dialects, viz., *miro*, *mlo* "my," *tiro/kiro*, *tlo/klo* "your." The forms *mo* and *to*, which are very common and are used in the text of RG are not mentioned in the exposition, nor are the reflexive possessive forms *piro*, *plo*, *po*, etc., "one's own." (We should note here that the forms in *-o* are masc. nom. sg. and that pronominal adjectives inflect like other agreeing adjectives.) This is another area that will require systematic literary codification.

2.3 Pronouns. Aside from the possessive pronominal adjectives just mentioned, there are three points connected with personal and demonstrative pronouns in RG that will need to be clarified and elucidated in the codification of Literary Romani.

The reflexive pronoun is given only in its full singular form *pes* in the exposition, although the reduced form *pe* is widely used to form various types of intransitives, as exemplified in the text, e.g., *ka vakjerel pe*

⁷ This by no means exhausts the possibilities, but in the context of our purposes it suffices.

"it will get discussed" (v. section 2.4). The plural form *pen*, which occurs in some Balkan and other Romani dialects (Ventcel' and Cerenkov 1976:311), is not used in RG nor, presumably, in the dialects on which it is based, e.g., *te garavan p̄es* "in order to defend themselves."

The third person pronouns occur with and without prothetic *v-* or *y-*, e.g., *ov*, *vov*, *yov* "he," etc. In general, RG uses the nonprothetic forms, although 3 sg. masc. and 3 pl. forms with prothetic *v-* (*vov*, *von*) are mentioned. In both the exposition of declension and the text, the initial *o* of the oblique cases (long forms) is written separately, as if it were a definite article, e.g., masc. loc. *o leste*, fem. loc. *o late*, pl. loc. *o lende*. (Sometimes the pl. nom. form is written *o la*, other times *ola*). There are both diachronic and synchronic arguments against this practice. Diachronically, the *o-* is part of the pronominal root, just as it is for demonstratives such as *okova* "that" (v. Sampson 1968:161). Synchronically, the definite article changes according to gender, number, and case (v. sections 2.2 and 2.4), but the *o-* of the oblique forms of the third person pronouns never changes, although it is omissible, e.g., *o argatija dikhle o le* "the field hands saw him," but also *nae ko te kjedol len* "there is no one to gather them." As was mentioned in section 1.7, there are still issues of word division that remain to be settled by a Literary Romani orthographic convention, and this is one of the.

Throughout the text, RG uses a pronoun, *kova*, as a relativizer, e.g., *manusa kola corena* "people who steal," *o manusa maskar kolende vestinena* "the people among whom they live." This pronoun is not mentioned in the section on pronominal morphology nor in that on relative clauses, and it is thus a phenomenon in need of further clarification.

2.4 Definite article. The dialects of RG belong among those that do not have *l-* in the oblique cases of the definite article (cf., Sampson 1968:152; Ventcel' and Cerenkov 1976:329-30). The dialects of RG all have *o* for the masc. nom. sg., *i* for the fem. nom. sg., and *e* for all oblique forms. In the nom. pl., both *o* and *e* are permissible, the use of *o* being more characteristic of the masc. pl. of the Arlija dialect. Here again we have the type of variation that will be of concern to the future discussion of codification.

2.5 Verbs. The Romani verb expresses a variety of traditional Indo-European categories and displays distinct complexities in its inflection. A normative grammar of Literary Romani will have a considerable task before it in unifying the diversity of dialectal variation in the Romani verbal system. The treatment of this system in RG brings to light facts which have not been observed in the various dialect descriptions, but there is also a considerable amount of material left untreated that will be of concern to codifiers. Here we will discuss some of the most important problems for the analysis of this system raised by RG or implied by its textual data.

The discussion of intransitive verbs mentions four types: 1) simple intransitives, e.g., *sovel* "he sleeps," 2) reflexives, 3) impersonals, and 4) passives. For reflexives, RG mentions only constructions with invariant *pes*, e.g., *me uravava pes* "I get dressed." Elsewhere in the text, however, RG also employs examples using reflexives with agreeing personal pronouns, e.g., *me uravgjum man* "I got dressed." The distinction here is one of dialectal usage (cf., Ventcel' and Cerenkov 1976:310). We can also mention here the ethical dative construction, which is not discussed in the exposition of grammar in RG. Although this construction is not, strictly speaking, reflexive, it does involve the relationship of the subject to the verb. The examples which occur in the text of RG all involve agreeing pronouns, as opposed to the Macedonian construction involving the invariant reflexive, which is given here for comparative purposes, e.g., *muk ma te dzav mange khere* (Mac. *Pusti me da si odam doma*) "Let me go home," *dza tuke* (Mac. *odi si*) "go away!"

RG gives examples of a number of different types of impersonal constructions without any further classification. We can identify the following types: derived (or "subjectless") impersonals, e.g., *rakjola* "it gets dark" (*rat* "night"); "oblique subject" impersonals, e.g., *ladz i ma* "I am ashamed" (literally "shame is me"), *pharpa i mange* "it's hard for me" (literally "difficulty is to me"), *sovel pes mange* "I'm sleepy; I feel like sleeping" (literally "it sleeps itself to me," cf., Mac. *mi se spie*). It should be noted here that RG also cites the derived personal intransitive *ladzovava* "I am ashamed," which, due to its marking for person does not belong in this category. The possessive construction with the third person of "be" and the accusative of the possessor, e.g., *[i]si man* "I have" can also be included here, but the negative of this construction using the invariant form *nae* "it isn't; it doesn't exist," e.g., *nae man* "I don't have" belongs in the next category.

For invariant nonfinite deverbatives, RG only mentions *valjani* "it is necessary," e.g., *valjani te khosav* "I have to shave," although *saj* "it is possible" and *nasti* "it is impossible" also belong here, as does the construction with *nae* just cited.

RG does not mention passive constructions in its exposition of intransitivity, although such constructions do occur in the text, e.g., *ka pokjinela pes tuke* "you will be paid" (literally "it will be paid to you"), *na ili agjaar sukar sikljovkjerdi* "it [history, fem.] has not been sufficiently well studied" (cf., also 2.3).

With verbs, as with other parts of speech, RG classifies derivatives by base rather than by suffix (cf., 2.1). The three main types illustrated by the examples given in RG are those using the verb *kerel* "do," e.g., *dromkjerav* "I travel" (*drom* "road"), those using *ovel* "be, become," e.g., *parnjovel* "whiten" (*parno* "white"), and unsuffixed denominals, e.g., *thagarel* "rule, be king over" (*thagar* "king"). The main suffix for adapting new loan verbs in the dialects of RG, not mentioned in this section, is *-in*, e.g., *mislinel* "think" (Mac. *misli*), *konstatirinel* "state, substantiate" (Mac. *konstatira*). Another suffixed verb commonly used in derivation is *del* "give," e.g., *cumidav* "I kiss" (*cumi* "kiss").

The treatment of aspect in the various dialects of Romani shows a considerable amount of variation. There are two major types of aspect to be considered: borrowed and native.

Under the influence of Slavic and Germanic languages, some dialects of Romani have adopted a perfective/imperfective-like aspectual opposition marked by prefixation as in the source languages. According to RG, the dialects upon which it is based do not have such an opposition. Nonetheless, there does appear to be

some borrowing of Slavic aspectual prefixes, although apparently with lexical-derivational functions, as exemplified by usage in the text, e.g., *vakerel* "say," *povakerel* "express." The section on participles in RG also claims that an aspectual opposition similar to that of the Macedonian perfective/imperfective can be rendered by the use of auxiliary verb suffixes, e.g., *gjlavdo* "sung" (Mac. *pean*), *gjlavkjerdo* "sung to the end" (Mac. *ispean*). This remains a task for further investigation. (Cf., also below on the present tense suffix *-a*.)

Romani does have an inherited aspectual opposition, viz., aorist/imperfect, a distinction which is also expressed in Macedonian as well as in other Balkan languages.⁸ The treatment of the preterit tense forms in RG brings to light constructions which have not been mentioned in other dialect descriptions, but it also omits forms which are still in use among the Roms of Macedonia. The three preterit forms discussed in RG are the synthetic aorist, e.g., *keravgjum* "I did," and two analytic preterits using the third person aorist of "be" (*sine*) as an invariant auxiliary verbal particle after the conjugated forms of the present and the synthetic aorist, e.g., *keravava sine*, *keravgjum sine*. There is no clear distinction made among these forms in the exposition in RG, and the best that can be inferred from the presentation is that *present + sine* corresponds to the Macedonian imperfect (including the perfect and pluperfect based on the imperfect stem) while *synthetic aorist* and *aorist + sine* can correspond to any Macedonian preterit form, viz., imperfect, aorist, perfect, or pluperfect. There are two problems with this analysis.

First of all, the exposition of verbal tense forms does not mention the synthetic imperfect (formed by adding *-s* to the long form of the present, e.g., *keravas* "I did/was doing/used to do," cf., below). It may well be that analytic preterits are replacing the older synthetic ones, especially in the speech of younger speakers, but the synthetic imperfect is still alive and well in the Romani dialects of Macedonia and should thus at least be mentioned in RG. Also, as the synthetic imperfect is still functioning in many other Romani dialects, it will have to be taken into consideration in the codification of the literary language.

The second problem concerns the analysis of the meanings of the various preterit forms. In its manner of presentation by comparison with Macedonian, RG gives the impression that the Romani preterit tenses do not form a system with contrasting basic meanings but rather that they are somehow all interchangeable. It is to be expected that the Romani and Macedonian preterit systems are not completely isomorphic and that as a result of this fact a given Romani preterit form will correspond to one or another Macedonian preterit form, depending on the contextual variant meaning expressed. Nevertheless, it is also to be expected that the preterit system of the dialects of RG, despite the fact that they may be in a state of flux due to competition between synthetic and analytic forms, will be governed by its own rules of usage which are statable in terms of basic and contextual variant meanings. It is the identification and systematization of these rules that will be a major task of the linguists involved in the codification of Literary Romani.

Another problem concerns the use of the so-called long forms in the present tense. The dialects of RG have the possibility of omitting the final *-a* of all the person markers of the present tense, e.g., *kerava* ↔ *kerav* "I do," *keresa/kereja/kerea* ↔ *keres/kere* "you do," *kerela* ↔ *kerel* "s/he, it does," etc. (In its description of the 2 sg/1 pl marker, RG mentions only Arlija *-[j]a* but not Dzambaz *-sa*; v. section 1.3.) According to RG, the short forms tend to occur in subordination to the subjunctive particle *te* and as the equivalent of the Macedonian perfective future when used with the future marker *ka*. In actual usage in the text, however, there is considerable variation in the presence and absence of this final *-a*, and in some dialects it has been lost altogether (Ventcel' and Cerenkov 1976:318; cf., Sampson 1968:189). This all raises the very important question of assigning grammatical function to linguistic variation. For example, during the early stages of the codification of Literary Albanian at the end of the nineteenth century, Faik Konitza suggested that the Geg (North Albanian) indefinite article *nji* and the Tosk (South Albanian) indefinite article *një* both be adopted by the literary language. He suggested that *nji* be used with feminine nouns and that *një* be used with masculine ones despite the fact that such a grammatical distinction is not carried by the indefinite article in either of the two dialects (Byron 1976:50). His suggestion was not accepted and Literary Albanian was thus spared an artificial grammatical distinction that did not exist in any spoken form of the language. In the formation of Literary Bulgarian, on the other hand, the two major dialectal forms of the definite article, viz., *-ut* and *-ŭ*, were both adopted, and the latter was assigned the function of distinguishing oblique noun phrases (cf., Puřev 1975:62-98). The codifiers of Literary Romani will be faced with similar problems in dealing with the retention and/or usage of short vs. long forms of the present tense.

With regard to stem vowels, there are two methods of analyzing the principal present tense conjugation patterns of Romani verbs: 1) classification according to the vowel preceding the consonant of the third person singular marker, viz., *e*-conjugation, e.g., *perel* "fall," *a*-conjugation, e.g., *dzal* "go", and *o*-conjugation, e.g., *tergjol* "stand"; 2) treatment of the *e* in, e.g., *perel* as part of the person marker which truncates after *a*-stem verbs like *dzal* (root *dza-*) and *hal*"eat" (root *ha-*) and which elides in verbs formed with *ovel* "be, become," since *tergjol* can also occur as *tergjoval* (cf., Ventcel' and Cerenkov 1976:315-17 vs. Sampson 1968:188-90). While the second method is better suited to a sophisticated linguistic analysis of the structure of Romani, the former fits better into the framework of traditional normative and pedagogical grammar. As might be expected, RG opts for the first type of analysis. In doing so, however, it completely omits the third type of conjugation (*tergjovav*"I stand," *tergjol* ↔ *tergjoval* "s/he it, stands," *tergjovena* ↔ *tergjona* 'all/they stand," etc.) from its exposition and so only cites the *e*-conjugation (*perel*) and *a*-conjugation (*dzal*).

⁸ The exact meanings of the categories *aorist* and *imperfect* vary from language to language, but the general aspectual concepts to which they relate are of the same basic nature (cf., Friedman 1983). The meanings of these categories for Romani have yet to be precisely defined.

types will be an important task for codification.

Standardization of the synthetic aorist is also a problem, since it is formed on the basis of the verbal participle, which shows considerable variation in its derivation even within a single dialect of Romani (v. Sampson 1968:196-203), although there are some general rules that can be formulated on the basis of the final consonant or vowel of the verbal stem (v. Ventcel' and Cerenkov 1976:316-17). The treatment of these phenomena in RG represents one of those areas most in need of additional work in the codification of Literary Romani.

In the exposition of the synthetic aorist, RG gives one paradigm (1 sg. *gjlavgjum* "I sang" [participle *gjlavdo*]) and cites an example sentence that employs another type (*peljum* "I fell" [participle *pelo*, irregularly formed from *perel*]), but it does not formulate any rules for predicting the form of the synthetic aorist or point out its morphological connection to the participle. In its section on the verbal adjective (i.e., participle), RG cites *-do* as the major formant and then cites the examples *gjlavel-gjlavdo* "sing-sung," *legarel-legardo* "carry-carried." Aside from citing as participial endings forms which could be analyzed as auxiliary verbal participles in *-do*, e.g., *-kjerdo* (from *kerel*), RG cites the suffix *-lo* with the example *dikhlo* "seen" and *-ndo* with the example *dendo* "given" (from *dikhel* and *del*, respectively). As can be seen from this summary, the presentation in RG does not give any idea as to whether the aorist and participles bear any systematic relationship to one another or to the verbal stems on which they are based. The presentation in Sampson (1968:196-203) gives far more data but a totally confusing picture. Ventcel' and Cerenkov (1976:316-17) set up rules which are, for the most part, applicable to the dialects of RG. If the verbal stem ends in *-r*, *-l*, *-n*, or *-v* the participial formant will be *-d-*; for most stems in *-d* the formant will be *-in-*; elsewhere the formant will be *-l-*; and there are approximately thirteen "irregular" verbs left over, e.g., *dzal* "go," masc. ptl. *gelo*, 1 sg. aor. *geljum*; *merel* "die," masc. ptl. *mulo*, 1 sg. aor. *muljum*, etc. For the dialects of RG, we need only add that *-in-* may elide to *-n-*, and that this *-n-* may itself be treated as stem final and hence have a suffixed *-d-* either in the aorist or in both the participle and the aorist. This explains not only the form *dendo* from *del* but also, e.g., *cidel* "drag," masc. ptl. *cidino* "dragged," 1 sg. aor. *cingjum* (based on a jotation of **cidindo*). Stems in *-nd*, however, will behave like other nonsonorant stems and take the formant *-l-* with the subsequent loss of the *-d*, e.g., *phanfel* "tie," masc. ptl. *phanlo* "tied" 1sg. aor. *phanljum*. The complete formulation of rules and exceptions, whether for the dialects of RG or for the future establishment of Literary Romani remains an important task for the future.

Another question in the use and form of the aorist that has yet to be addressed is that of the form of the third person singular aorist. The usual ending, which is illustrated by RG in its exposition, is *-a*, e.g., *gjlavgja* "s/he sang," *halja* "s/he ate." Certain verbs with the participle in *-l* (more rarely also stems in *-d*), however, have third singular aorist forms identical with those of the participle, which is to say that they agree with the gender of the subject like other adjectives, e.g., masc. *gelo* "he came," *mulo* "he died," fem. *geli*, *muli*.⁹ These facts are illustrated by the text of RG but are not discussed in its exposition and thus remain a problem for future codification.

The forms and uses of the verbs meaning "be" (*isi*[*si*, *i*] and *ovel*, which also means "become") especially the participle *ili*, etc., will require considerable further elucidation in the codification of the literary standard.

2.6 *Prepositions*. As a relatively late development in the Romani language, most prepositions are subject to a considerable amount of variation in both form and usage both from dialect to dialect and within a given dialect (cf., e.g., Sampson 1968:221-22; Ventcel' and Cerenkov 1976:227-28; Kostov 1962:143-45). This is thus an area that will require a careful and thorough analysis in the codification of Literary Romani. In RG, the treatment of prepositions provides useful information, but we can make the following observations.

There is considerable variation in case government. In some instances, substantives behave differently from pronouns, e.g., *pali o nilaj* (nom.) "after the summer" but *pali mande* (loc.) "behind me." In other instances, the same preposition can change meaning with a change of case (cf., Ventcel' and Cerenkov 1976:328), e.g., *me dzava kji Bitola* (nom.) "I'm going to Bitola," but *kji makedoniakjere chibate* (loc.) "in the Macedonian language"; this differentiation is not consistent in RG, however, e.g., *kji romani chib* (nom.) "in the Romani language."

Aside from variation in the actual form of prepositions, e.g., *uzo/uzal* "according to," there is also variation with regard to independent and compound prepositions, e.g., *me tergjovava dzi* [+ *ko*] *duvari* "I am standing by [cf., next to] the wall."

A number of Romani prepositions elide with the definite article and are sometimes (but not in RG) spelled with an apostrophe or hyphen, e.g., *ko/k'o/k-o gav* "to the village," *ki/k'i/k-i diz* "to the town." The original shape of some of these prepositions can be inferred from case endings, which were originally these same words functioning as postpositions, e.g., *gaveski*, *dizjake*. In RG, only the preposition *po/pi* "along, on, etc.," is described as varying according to gender. It is not described as elision with the definite article, the oblique form *pe* is not mentioned, nor is any mention made of other prepositions affected by this phenomenon, although they are used in the text.

Under the influence of the grammatical structure of Macedonian, RG mentions simple case form uses that correspond to Macedonian prepositional phrases, e.g., *silestar* "from the cold" (Mac. *od stud*), *mareskje* "for bread" (Mac. *za leb*). These do not belong in a section on prepositions but rather in the section on case form

⁹ The third plural aorist ending is always identical with the plural adjective ending, viz., *-e*, e.g., *gjlavgje* "they sang," *gele* "they came" and so the problem in question only arises in the singular.

usage.

2.7 *Other uninflected parts of speech.* These present problems mainly on the basis of dialectal differentiation and can be treated lexically, e.g., the adverbs *akhal* and *agjaar* "thus."

3 *Syntax.* The treatment of this subject in RG represents the most conservative taxonomic school of thought. Phenomena connected with word order, reduplication of objects, case government of prepositions, etc., are virtually unmentioned. As in most languages, so in Romani the description of syntax is an area where much work remains to be done. Given the fact that syntax is an area of linguistics where debate on fundamental principles is still in progress, its treatment in the codification of Literary Romani will depend largely on the theoretical orientation of the codifiers.

4 *Dialect choice.* As was indicated in the first paragraph, RG is based on the Arlija dialect of Skopje and to a lesser extent on Dzambaz, followed by Gurbet and Burgudzi. When a choice of forms arises, e.g., the adverb *akhal/agjaar* "thus" (Arlija/Dzambaz) or the accusative pronouns *man, tut/ma, tu* "me, you," RG frequently uses the Arlija form in its Romani text and the Dzambaz form in the Romani examples in its Macedonian text. The decision regarding which forms to include as coexisting and which forms to exclude from Literary Romani will be one of the major tasks of codification.

5 *Conclusion.* Having examined some specific problems connected with the standardization of a Romani literary language raised by the text of RG, it will be appropriate for us to conclude by making the following two points. While the codification of Literary Romani involves a variety of language specific data, the basic issues and methods at stake are the same as those of the codification of literary languages in general in modern times. A major step in the direction of codification has been taken in the publication of RG, which will stand as a landmark in the development of Literary Romani.

REFERENCES CITED

- Byron, Janet. 1976. "Faik Konitza dhe gjuha letrare shqipe." *Faik Konitza 1876-1976*. Edited by E. A. Licho, pp. 49-51. New York: Vatra.
- Friedman, Victor A. 1983. "Grammatical categories and a comparative Balkan grammar." *Ziele und Wege der Balkanlinguistik*. Edited by Norbert Reiter, pp. 81-98. Wiesbaden: Harrassowitz.
- Hancock, Ian. 1975. *Problems in the creation of a standard dialect of Roman s.* (Social Sciences Research Council Working Papers in Sociolinguistics No. 25.) Washington, D.C.: E.R.I.C.
- Jusuf, Saip and Krume Kepeski. 1980. *Romani gramatika - Romska gramatika*. Skopje: Nasa Kniga.
- Jusuf, Saip. 1981. "Amari chib." *Loli Phabaj* 1:6-12.
- Kenrick, Donald. 1981. "Romano alfabeto." *Loli Phabaj* 1:3-4.
- Kepeski, Krume. 1972. *Gramatika na makedonskiot literaturn jazik za ucilistata od II stepen*. Skopje: Prosvetno Delo. (Third edition, first published in 1950; First edition, 1946.)
- Kochanowski, Jan. 1963. *Sypsy studies* (two parts). New Delhi: International Academy of Indian Culture.
- Kostallari, Androkli et al. 1974. *Drejtshkrimi i gjuh s shqipe*. Prishtina: Instituti Albanologjik i Prishtin s.
- Kostov, Kiril. 1962. "Aus der Syntax der Zigeunersprache Bulgariens." *Balkansko ezikoznanie* 4:131-46.
- Puřvev, Xristo. 1975. *Ocerk na istorija na bulgarska gramatika*. Sofia: Nauka.
- Puxon, Grattan. 1978/9. "O Romano Kongresso 2: The Second World Romani Congress." *Roma* 4,2/3:33-104.
- Ristovski, Blaze, ed. 1974. *Gjorjgija M. Pulevski: Odbrani stranici*. Skopje: Makedonska Kniga.
- Sampson, John. 1968. *The dialect of the Gypsies of Wales*. Oxford: Oxford University. (First published 1926).
- Tosev, Krum et al. 1970. *Pravopis na makedonskiot literaturn jazik*. Shopje: Prosvetno Delo.
- Veřtel', T. V. and L. N. Cerenkov. 1976. "Dialekty ciganskogo jazyka." *Jazyki Azii i Afriki I*. General editor N. I. Konrad, pp. 283-332. Moscow: Nauka.
- Vogt, Hans. 1971. *Grammaire de la langue g orgienne*. Oslo: Universitetsforlaget.