

Christine Hands
2011

Dialect classification: a study of Romani numerals

Introduction

Romani is a minority language which does not have a specific geographical territory; rather, it is spoken throughout Europe by groups which are traditionally considered to be peripatetic communities. The language is believed to have originated in India as a branch of the Indo-Aryan languages; it later separated from Indo-Aryan and formed different dialects as a result of the migration of Rom from India to Europe (Matras 2002).

Boretzky (2007) explains that the Rom arrived in the Balkan region of Europe from the 11th century onwards in a number of waves. The different groups, sometimes referred to as 'clans', then dispersed throughout Europe and now speak a diverse set of Romani dialects. Despite such variation between Romani dialects due to the influence of surrounding languages, the dialects are still mutually intelligible and said to be united linguistically as a single language (Matras 2008). Although linguists "agree on the fact that the Romani dialects can be classified into a number of dialect groups (families), [they disagree] on how these groups have come about" (Boretzky 2007: 314), for example, Boretzky (2007) and Matras (2002, 2005) have different standpoints on this matter; these will be discussed below in a review of the dialect classification schemes which they propose.

The main focus of this paper will be upon the numeral systems of a selection of Romani dialects. Although there has been some discussion of Romani numeral systems in the past (e.g. Bakker 2001, Elšík & Matras 2006), it will be interesting to find out if they can be related to and support dialect classification schemes, specifically the dialect classification scheme that has been proposed by Matras (2002, 2005). In order to do this, a sample of dialects will be examined to find out how they pattern in regards to shared inherited numerals, borrowed numerals, and how dialects form numerals internally using existing numerals in the system.

Early dialect classification schemes

The diverse dialects which comprise the Romani language have been a topic of historical and structural interest for scholars, for example, Pott (1844-5 cited in Matras 2002) conducted a comparative study of Romani dialects and highlighted the pre-European loan vocabulary which exists in the Romani lexicon. Building on the work of Pott, Miklosich (1872-80 cited in Bakker & Matras 1997 and Matras 2002) proposed a pioneering dialect classification scheme which divided the Romani dialects into 13 groups based on an examination of the different layers of lexical borrowings in Romani (e.g. from Iranian, Armenian and Greek). From these lexical borrowings he was able to reconstruct the routes taken by the Rom when they migrated from India and travelled to and within Europe. The large number of Greek

This paper was written as part of coursework requirements for an undergraduate module on Romani Linguistics taught by Professor Yaron Matras (School of Languages, Linguistics and Cultures, University of Manchester, 2010-2011).

loans “led Miklosich to conclude that a Greek-speaking area had been the European homeland of all Gypsies before their dispersion across the continent” (Bakker & Matras 1997: ix).

Some years later, another influential attempt to classify the Romani dialects was made by Gilliat-Smith (1915 cited in Bakker & Matras 1997), who distinguished between Romani dialects in northern Bulgaria according to a classification as either ‘Vlax’ or ‘non-Vlax’. This classification was later extended to dialects which had migrated from Wallachia, but which shared particular structural features with Vlax dialects, features which were often the influence of Romanian.

Modern dialect classification schemes

The geographical diffusion model proposed by Matras (2002, 2005) will be focused upon when analysing the numeral systems of different Romani dialects. This dialect classification scheme arranges the Romani dialects into the following groups based on different isoglosses: Northwestern, Northeastern, Central, Vlax, and Balkan. These groups are generally accepted in the literature, although there is some variation, for example, Boretzky (2007) divides the Balkan group into two and also divides Central and Vlax into north and south groups. It is important to note that these groups are “based on impressions of a series of shared features” (Matras 2005: 10), however, they are also somewhat arbitrary.

An additional difference between the dialect classification schemes proposed by Matras (2002, 2005) and Boretzky (2007) is found in the explanations they offer as to how the different dialect groups emerged. Matras (2002) suggests that the Romani dialects began to diversify linguistically from the 15th century onwards when the different clans arrived at the regions in which they would settle. Innovations then spread from dialect to dialect by geographical diffusion; consequently, they form a geographical continuum. This explanation is based on the idea that “Early Romani was the uniform stem from which from which individual dialect branches descended” (Matras 2002: 215). In contrast, Boretzky (2007: 319) does *not* propose that Early Romani was a uniform stem; instead he suggests that “Romani since early times displayed a dialectal restructuring that became stronger and stronger by the innovations taking place on the way to Central and Northern Europe”. By this view, the Romani dialects do not form a geographical continuum.

The geographical diffusion model operates on the principle that changes and innovations spread from dialects “gradually over time and space” into neighbouring dialects (Matras 2002: 14). Matras (2002: 236) explains that this theory can account for a number of diagnostic isoglosses which represent differences between dialects and that a dialect can be classified according to its “participation in a cluster of isoglosses”. These isoglosses are believed to have been formed as a result of the spread of innovations from three different centres of diffusion: south-eastern Europe, western-central Europe, and Vlax. However, it is also pointed out that not all innovations spread for these three centres; some innovations are the outcome of language contact which spread and are restricted to specific regions (Matras 2002).

Numeral systems

In the present paper, the numeral system is studied in order to find out what cross-dialectal variation there is in Romani and also to find out what any variation may indicate in relation to dialect classification. Specifically, I will be looking at the cardinal numerals from 1-10, 11-19, and the tens numerals (multiples of 10) from 20-90.

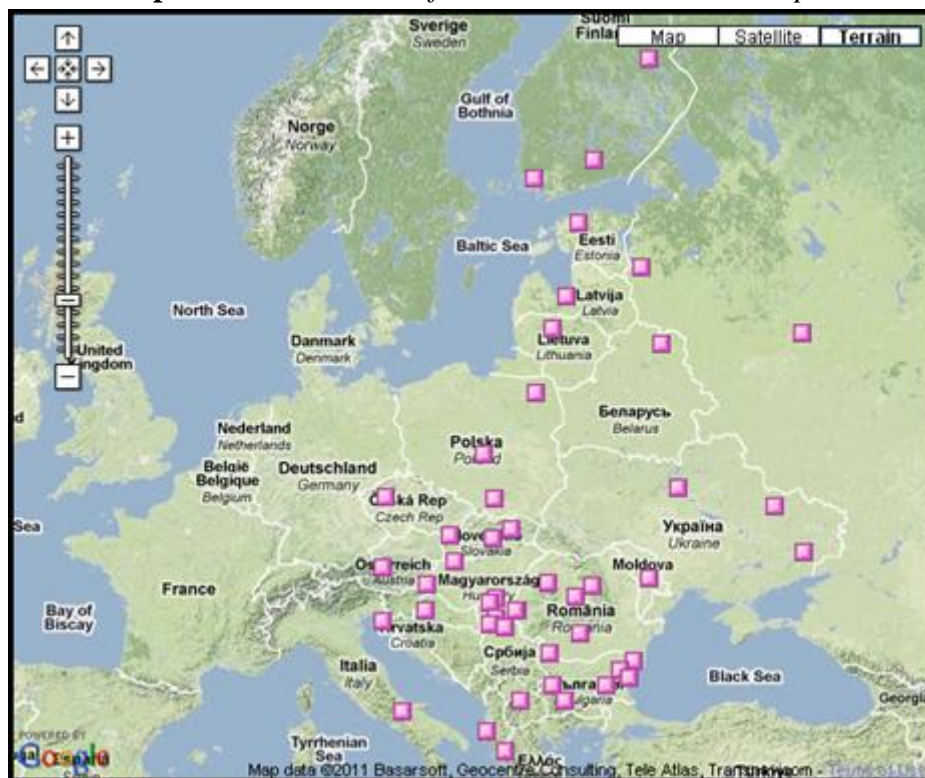
It is important to note that a number of general tendencies can be observed across the numeral systems of different languages and these tendencies will be taken into account when investigating the numeral system of Romani. For example, Hurford (1987: 8) states that:

A number x is named by an expression whose constituents are the names of the numbers y and z .

This general tendency is stated in regards to numbers such as 19 which are formed in different languages “in an exactly parallel way” (Hurford 1987: 13), i.e. from the combination of the numeral forms for 9 and 10.

A sample of 47 Romani dialects has been taken from the Romani Morpho-Syntax (RMS) Database (see Matras & Elšík n.d.); this database has been the sole source of information on Romani numerals for this investigation. The dialects have a wide geographical spread so that the sample is representative of diversity (see map 1). A full list of the dialects studied can be found in appendix 1.

Map 1 *The distribution of Romani dialects in the sample*



Numerals 1-10

Little variation was found in the set of numerals from 1-10; from the sample it would appear that these numerals are remarkably consistent across dialects. Generally speaking, dialects were found to preserve inherited forms from Indo-Aryan for the numbers 1-6 and 10; they also preserve Early Romani loans from Greek for the numbers 7, 8 and 9 (*efta*, *oxto*, *enja*). Greek numerals such as these demonstrate the influence which the Greek language had upon the Romani lexicon and support Miklosich's (1872-80 cited in Bakker & Matras 1997) claim that the Rom had a prolonged period of contact with a Greek speaking area before their dispersal throughout Europe. Any variation found is mainly phonological; this can be illustrated by the different numeral forms for 9 which were found, for example, *enja*, *inja*, *īja*, *ja*, and *ənje*, amongst others. The observations presented here are consistent with Bakker (2001), who also notes the homogeneous nature of this numeral set.

The examination of numerals from 1-10 has revealed that this set has been largely resistant to change and their homogeneity could be taken as support for Matras' (2002, 2005) claim that when the Rom came to Europe they spoke a largely uniform language. However, it is difficult to draw any further conclusions from the dialects which share these numerals in relation to dialect classification.

Numerals 11-19

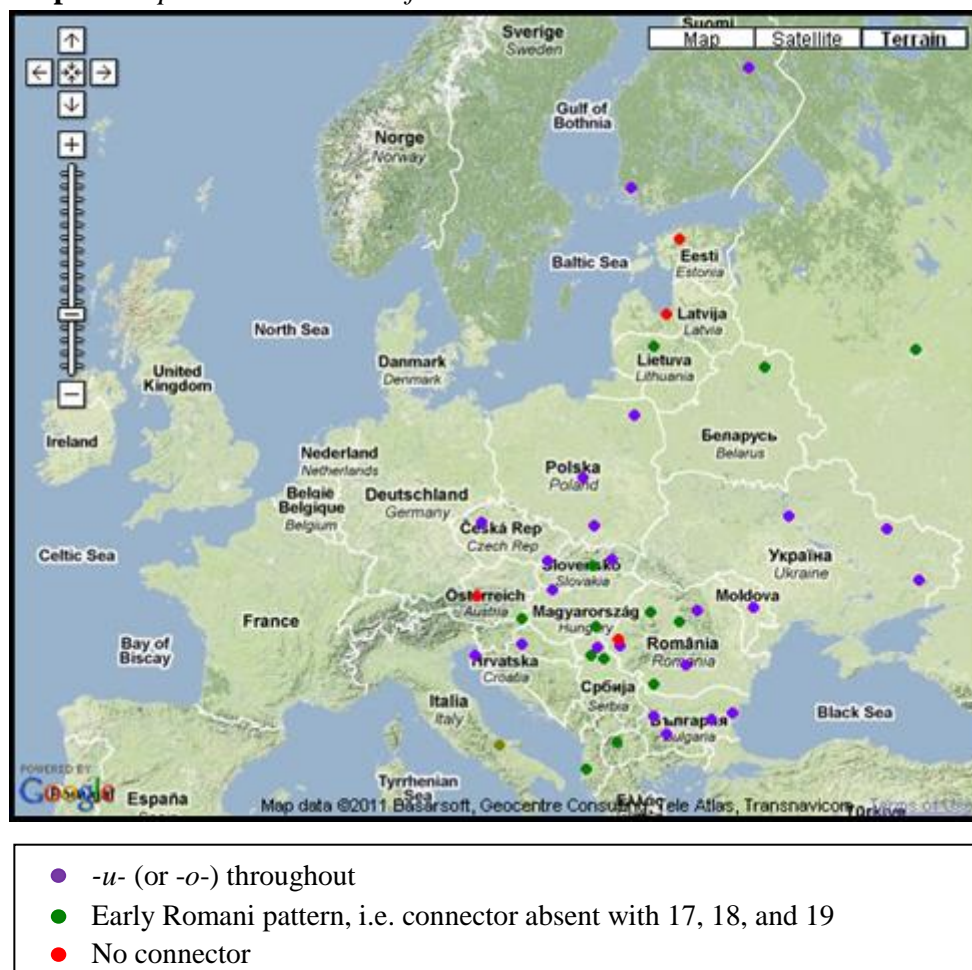
An examination of the set of numerals from 11-19 revealed that they are predominantly formed by addition of *deš* '10' and a unit numeral via a connector, typically *-u-* (or *-o-*) 'and'. According to Elšík & Matras (2006), the connector *-taj-* is also used in some dialects, however, this was only attested in 1 dialect from the 47 in the sample; this was the Molise dialect (taking the form *-ta-*).

Elšík & Matras (2006: 164-5) explain that "Early Romani connected the indigenous unit numerals '1' through '6' by means of an overt connector, but used no connector with Greek-derived unit numerals '7' through '9' (e.g. *deš-u-šov* '16' vs. *deš-efta* '17')". This pattern was found to be preserved in 13 dialects from the sample (e.g. Ursari and Gurvari). Those dialects which do not follow this pattern have either the predominant pattern mentioned previously, whereby the use of a connector has been generalised to all numerals from 11-19, or else they have generalised the null marker used for 17, 18, and 19 in Early Romani to the lower numerals in this set. However, the latter pattern was attested only in Austrian and Romanian Sinti and would therefore seem to be a rarer development, although the use of a connector is almost completely absent in Čuxny and Lotfitka too, for example, in Čuxny the connector *-u-* only remains in use with the number 11. Conversely, there are also a few dialects which would appear to have begun generalising the use of a connector to all numerals in the set but in which this generalisation has not yet been completed, for example, in Curjarja Arilje spoken in Croatia the only number which does not feature the connector *-u-* is 17.

Map 2 illustrates the distribution of the different patterns; however, not all dialects in the sample could be represented due to incomplete data sets on the RMS Database. Čuxny and

Lotfitka have both been marked with a red dot for the absence of a connector in numerals 11-19 as it is assumed here that this pattern will eventually become fully generalised in these dialects.

Map 2 The presence/absence of a connector in the cardinal numerals 11-19



Dialects which retain the Early Romani pattern for numerals 11-19 (green dots on Map 2) seem to be found in an area which spans over parts of what Matras (2002, 2005) has referred to as the Balkan, Vlax, and Central dialect regions. However, there are also a number of dialects interspersed with these which have generalised the more common and more geographically widespread pattern, whereby a connector is used with all numerals from 11-19 (purple dots). Therefore, it does not seem plausible to suggest that the pattern a dialect exhibits is the result of geographical diffusion since there does not appear to be any clear patterning, rather, as Elšík & Matras (2006: 165) suggest, it is more sensible to conclude that the different patterns exhibited have developed in individual dialects “irrespective of their origin”.

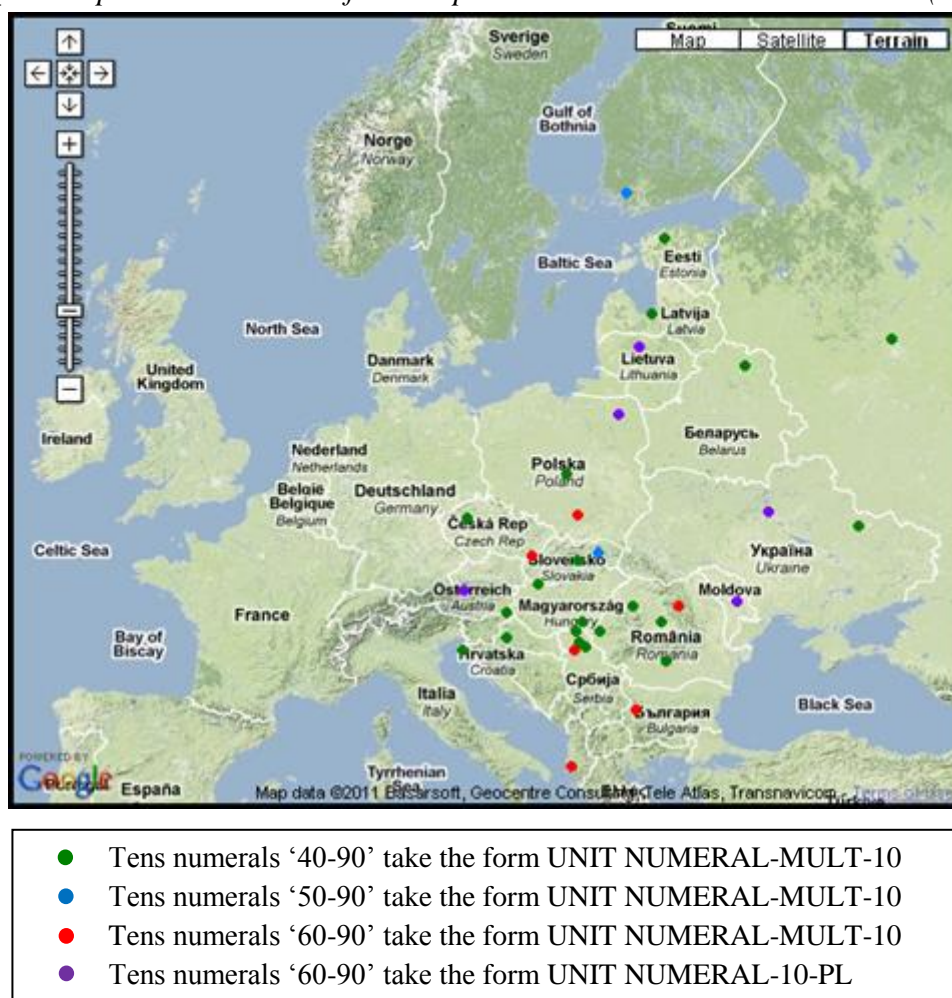
The tens numerals (20, 30, 40 ... 90)

The numeral *biš* ‘20’ is found in the majority of Romani dialects; according to Elšík & Matras (2006: 168) it is an “underived indigenous” form. The numeral *tranda* ‘30’ which is

inherited from Greek is also maintained in most Romani dialects, whereas the numeral forms for 40 and 50 show more variation; some dialects use the Greek forms *saranda* and *penda*, respectively, as in Šušuwaje and Sofia Erli, others feature the use of internally formed compounds either with or without a multiplicative connector (Elšik & Matras 2006).

Almost all dialects in the sample construct the numeral forms for 60-90 using a combination of lower numerals in the system. The predominant pattern is to combine the relevant unit numeral with *deš* ‘10’ via a multiplicative connector, for example the numeral *eftavardeš* ‘70’ is formed from *efta* ‘7’, the multiplicative connector *-var-* (other attested variants found were *-val-* or *-va-*), and *deš*, literally meaning ‘7 times 10’. The numeral *eftadeša* ‘70’ (7-10-PL) is an alternative form found in some dialects and which does not feature a multiplicative connector, for example, this form was found in Polish Xaladytka (illustrated by the purple dot in the north east of Poland on Map 3). Once again, it was not possible to map all dialects in the sample due to insufficient data sets.

Map 3 *The presence/absence of a multiplicative connector in tens numerals (20-90)*



From Map 3 it would appear that there is a more coherent pattern for the formation of tens numerals cross-dialectally. Dialects which form the numerals 40-90 with a multiplicative connector (green dots) tend to occur in those dialects which are classified according to

Matras' (2002) classification scheme as Central, Vlax, and Northeastern (e.g. in Čuxny, Lotfitka, and Polish Xaladytka). The geographical distribution of the pattern found here is also noted by Elšik & Matras (2006: 171); furthermore, they also note the rarity of dialects which form only the numerals from 50-90 using a multiplicative connector and this can be seen in Map 3. Therefore, an analysis of the tens numerals reveals that dialects *do* seem to pattern in line with the dialect groups which have been proposed in the literature by the likes of Matras (2002, 2005). It cannot be said with certainty that such a distribution is the result of geographical diffusion, although it would certainly appear to be a possibility.

The influence of language contact on Romani numeral systems

Not all of the dialects studied using the RMS Database exhibited the same patterns as those which have been discussed above. For example, Xoraxani which is spoken in Bulgaria was found to have inherited Indic forms for 1-3 but has replaced all other numerals from the Early Romani system with Turkish loans (Schönig 1999, see appendix 2). In the Molise dialect of Italy the Indic numerals 1-6 are retained but the inherited Greek numerals 7-9 have been replaced with Italian numerals. Furthermore, Molise has Italian numeral forms for 2 and 6 which are used alongside the inherited Indic forms (see appendix 3). There is also the case of Russian Roma which has borrowed numerals due to contact with Russian, for example, numerals from 7-20 and also 70, 80, and 90 have been replaced with Russian numerals (Kochetov 2003, see appendix 4). The retention of Greek numerals for 30 and 40 in Russian Roma illustrates the point made by Elšik & Matras (2006: 162) which is that medium cardinal numerals are more likely to be retained than older borrowed cardinals.

The Rom are known to make a living through trade in the regions they are settled (Matras 2008) and as numerals figure prominently in trade domains it is not surprising that they are a borrowable element in Romani. Matras (2009: 58) explains that in situations of language contact “group B speakers will import into their own language word-forms acquired through interaction with group A in the relevant domain”. This can be applied to the situation of contact found between Romani speakers and the other languages with which they come into contact, for example, speakers of Romani dialects require access to trade domains which are dominated by speakers of a majority contact language (e.g. Turkish, Italian, and Russian in the cases of the three dialects discussed above) and this leads to the borrowing of numerals into the Romani lexicon. Contact between the Xoraxani dialect and Turkish must be particularly extensive since nearly the whole numeral system has been replaced.

According to Matras (2002), the geographical diffusion model is capable of accounting for dialects such as these which deviate from the patterns found in other Romani dialects. This is because they are dialects typically spoken in peripheral areas, and it is not uncommon for peripheral dialects to pattern differently from others.

Conclusion

It has been possible to relate the numeral systems of Romani dialects to the geographical diffusion model proposed by Matras (2002) to a certain extent. The numerals which displayed the most significant patterning in regards to this dialect classification scheme were

the tens numerals from 40-90 as there was found to be a more coherent distribution of dialects which employ compound forms using a multiplicative connector for these numerals; this was in an area comprised of a number of Central and Vlax dialects, in addition, some dialects in the northeast were found to form these numerals in the same way. On the other hand, the formation of numerals from 11-19 either *with* or *without* a connector did not reveal a consistent clustering of dialects which shared numeral forms and so geographical diffusion could not provide a plausible explanation for the distribution of different numeral patterns. In this case it is more sensible to assume that the pattern of formation exhibited by a dialect has no relation to their origin, rather, the developments are more likely to have occurred in dialects individually.

It has become evident from this study of Romani numerals that language contact is an important factor influencing the Romani numeral system. Due to the fact that Romani is a minority language that comes into contact with different majority languages (depending on the region that a group has settled), this leads to the borrowing of words into the Romani lexicon and so speakers are generally bilingual (Matras 2008); the importance of the trade domain in Romani culture means that the numeral system is often subject to the influence of borrowed forms.

References

- Bakker, Peter. 2001. Typology of Romani numerals. *Sprachtypologie und Universalienforschung* 54(2), 91-107.
- Bakker, Peter & Yaron Matras. 1997. Introduction. In Yaron Matras, Peter Bakker & Hristo Kyuchukov (eds.), *The typology and dialectology of Romani*, vii-xxx. Philadelphia: John Benjamins.
- Boretzky, Norbert. 2007. The differentiation of the Romani dialects. *Sprachtypologie und Universalienforschung* 60(4), 314-336.
- Elšík, Viktor & Yaron Matras. 2006. *Markedness and language change: The Romani sample*. Berlin/New York: Mouton de Gruyter.
- Hurford, James R. 1987. *Language and number: The emergence of a cognitive system*. Oxford: Basil Blackwell Ltd.
- Kochetov, Alexei. 2003. *Numeral systems of the world's languages: Russian*. [Website]. Available from: <http://lingweb.eva.mpg.de/numeral/Russian.htm>. Accessed on: 27/01/2011.
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2005. The classification of Romani dialects: A geographic-historical perspective, 7-26. In Barbara Schrammel, Dieter W. Halwachs & Gerd Ambrosch (eds.), *General and Applied Linguistics*. Munich: Lincom Europa.
- Matras, Yaron. 2008. *Notes on Romani culture*. Course materials for Romani Linguistics, 2010/2011. Available from: <http://blackboard.manchester.ac.uk>.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Matras, Yaron & Viktor Elšík. n.d. *Romani Morpho-Syntax Database*. [Computer software].

Available from: <http://romani.humanities.manchester.ac.uk/rms/>.

Schönig, Claus. 1999. *Numeral systems of the world's languages: Turkish*. [Website].

Available from: <http://lingweb.eva.mpg.de/numeral/Turkish.htm>. Accessed on: 28/12/2010.

APPENDIX 1

Name of dialect	Location	RMS Database code
1. Russian Roma	Russia	RUS-003
2. Lovari Čekeši	Russia	RUS-005
3. East Finnish Romani	Finland	FIN-002
4. West Finnish Romani	Finland	FIN-005
5. East Finnish Romani	Finland	FIN-012
6. Čuxny	Estonia	EST-005
7. Lotfitka	Latvia	LV-006
8. Lithuanian Romani	Lithuania	LT-007
9. Polska Roma	Poland	PL-003
10. Bergitka	Poland	PL-007
11. Polish Xaladytka	Poland	PL-014
12. Polish Xaladytka	Belarus	PL-015
13. Czech Vlax	Czech Republic	CZ-001
14. East Slovak	Slovakia	SK-011
15. West Slovak	Slovakia	SK-016
16. Romungro	Slovakia	SK-027
17. Sinti	Austria	AT-001x
18. Lovari	Hungary	HU-004
19. Gurvari	Hungary	HU-007
20. Prekmurski	Slovenia	SLO-001
21. Kubanski Servy	Ukraine	UKR-008
22. Plasčuny	Ukraine	UKR-019
23. Gimpeny	Ukraine	UKR-020
24. Laješa/Kišinevcy	Moldova	MD-001
25. Ursari	Romania	RO-004
26. Kaldaraš	Romania	RO-008
27. Šušuwaje	Romania	RO-012
28. Kurtarare	Romania	RO-015
29. Sinti	Romania	RO-022
30. Maj Vlaši	Romania	RO-058
31. Kalderaš	Romania	RO-065
32. Gurbet-Rabešte	Serbia	YU-004
33. Bačkačjke	Serbia	YU-007
34. Arli	Serbia	YU-011
35. Lovari	Serbia	YU-015

36. Curjarja Arilje	Croatia	HR-002
37. Manuša Čurjarja	Croatia	HR-003
38. Thracian Kalajdži	Bulgaria	BG-007
39. Kalajdži	Bulgaria	BG-009
40. Goli Cigani	Bulgaria	BG-011
41. Rešitari/Čergari	Bulgaria	BG-012
42. Xoraxani	Bulgaria	BG-015
43. Sofia Erli	Bulgaria	BG-024
44. Kovački	Macedonia	MK-012
45. Mečkaria	Albania	AL-001
46. Romacilikanes	Greece	GR-002
47. Molise	Italy	IT-007

Appendix 2

Xoraxani, Bulgaria (RMS Database: BG-015)

Cardinals	Form	Origin
1	ek	Inherited
2	duj	Inherited
3	trin	Inherited
4	dört	Turkish
5	beš	Turkish
6	altə	Turkish
7	jedi	Turkish
8	sekiz	Turkish
9	dokuz	Turkish
10	on	Turkish
11	onbir	Turkish
12	oniki	Turkish
13		
14		
15	onbeš	Turkish
16	onaltə	Turkish
17	onjedi	Turkish
18	onsekiz	Turkish
19	ondokuz	Turkish
20	jirmi	Turkish
30	otuz	Turkish
40	kərk	Turkish
50	eŋi	Turkish
60	altməš	Turkish
70	jetmiš	Turkish
80	seksen	Turkish
90	doksan	Turkish

(cf. Matras & Elšík n.d. and Schönig 1999)

Appendix 3

Molise, Italy (RMS Database: IT-007)

Cardinals	Form	Origin
1	jek	Inherited
2	due	Italian
	du	Inherited
3	tri(n)	Inherited
4	štar	Inherited
5	panč	Inherited
6	sej	Italian
	šo	Inherited
7	sette	Italian
8	otto	Italian
9	nove	Italian
10		
11	deštajek	Inherited
12	deštadu	Inherited
13		
14		
15		
16		
17		
18		
19		
20	biš	Inherited
30	trijanda	Inherited
40		
50		
60		
70		
80		
90		

(cf. Matras & Elšík n.d.)

Appendix 4

Russian Roma, Russia (RMS Database: RUS-003)

Cardinals	Form	Origin
1	jek	Inherited
2	duj	Inherited
3	trin	Inherited

4	štar	Inherited
5	panč	Inherited
6	šov	Inherited
7	sjemj	Russian
8	vosimj	Russian
9	djevitj	Russian
10	djesitj	Russian
11	adinacytj	Russian
12	adinacytj	Russian
13		
14		
15	pitnacytj	Russian
16	šysnacytj	Russian
17	simnacytj	Russian
18	vasimnacytj	Russian
19	djevitnacytj	Russian
20	dvacytj	Russian
30	trijanda	Inherited
40	saranda	Inherited
50	pandeša	Inherited
60	šovdeša	Inherited
70	sjemdisjat	Russian
80	vosmdisjat	Russian
90	djevinosta	Russian

(cf. Matras & Elšík n.d. and Kochetov 2003)