

## Romani Orthographies<sup>1</sup>

Romani, the language of ‘the people known as ‘Gypsies’, Roma, or Roms’ (Matras 1999:481), is a predominantly oral language. Thus Sampson, in what Matras (2002:3) calls his ‘monumental’ 1926 work on Welsh Romani begins by referring to the language as ‘the speech of an unlettered people’ (Sampson 1926:3); and over seventy years later, Matras can still say the same thing: ‘Romani is primarily an oral language’ (Matras 2002:238); ‘Romani is typically and historically an oral language, and Gypsy culture rests entirely on oral traditions.’ (Matras 1999:482).

This in itself accurate assertion of the oral nature of Romani is to be distinguished from the ‘commonly repeated fallacy that Romani is not a written language’ (Hancock 1995:34). The Roma are still today a largely ‘unlettered people’ – very few are believed to be literate at all (Matras 1999:481), and of these only a small number, already literate in the national state language will also be able to read and write Romani (Matras 1999:482, 2002:251). Nonetheless such people do exist; and there is ‘limited tradition of literacy’ (Matras 2002:251) – limited, but stretching back for ‘over a century’ (Hancock 1995:34).

It is the purpose of this essay to give an overview of this history, examining some of the various and varying orthographies used to see how they differ. A broad distinction may be drawn between the orthographies of non-Romani scholars studying the language and those used by the Roma themselves. There have also been numerous attempts to create standardized orthographies for all or some dialects, with varying degrees of success. I will discuss each of these areas in turn.

First, however, the choice of writing system must be examined. Given that the Roma live in Europe, it is hardly surprising that Latin has been the script most commonly used to write Romani. It is not the only one, however. Romani texts were produced in the USSR in the 1920s and 1930s (Matras 1999:483, 2002:257; Friedman 2005:163), and for these the Cyrillic alphabet was used – (1) gives an example of a Cyrillic orthography for a Romani-Russian dictionary (Sergievsky & Barannikov 1938):

- (1) А а, Б б, В в, Г г, Г г, Д д, Е е, Ё ё, Ж ж, З з, И и, Й й, К к, Л л, М м, Н н, О о, П п, Р р, С с, Т т, У у, Ф ф, Х х, Ц ц, Ч ч, Ш ш, Ы ы, Ь ь, Э э, Ю ю, Я я.

Cyrillic remains in use for writing Romani today in Russia, Bulgaria and Serbia (Bakker & Kyuchukov 2000:90, 111), although in the latter two places the Latin alphabet has become more common (Matras 2004:6-7).

The Greek alphabet has also been used, though ‘very little seems to be written in Romani in Greece’ (Bakker & Kyuchukov 2000:90), and Arabic script was used for the first Romani periodical produced in Turkey in the 1920s (*ibid.*). The language has even occasionally been written in Devanagari, as a way of signalling its relation to Indian languages: Sampson included a Devanagari Romani text as a frontispiece (Sampson 1926:v; Bakker & Kyuchukov 2000:111), and the script is used today alongside Latin in the Romani Wikipedia. Latin

1. This paper was written as part of coursework requirements for a module on Romani Linguistics taught by Professor Yaron Matras at the School of Languages, Linguistics and Cultures, The University of Manchester, 2010-2011.

remains the script most widely used, however, and the remainder of this essay will examine some of the Latin orthographies used for Romani.

Long before the Roma ever wrote their own language, it had been written down by non-Romani scholars, so it is with them that the history of Romani orthography begins. The first written texts in Romani take the form of lists of words produced from the 16<sup>th</sup> century onwards: the earliest of these was a list of 13 sentences with an English translation collected by Andrew Borde and published in 1542 (Matras 2002:2; Bakker & Kyuchukov 2000:90). Numerous other such lists were published in various countries, and by the 18<sup>th</sup>-19<sup>th</sup> centuries much research was being carried out on the language (Matras 2002:2). Certain orthographic conventions therefore developed, ‘never conventionalized’ (Matras 1999:488), and thus not entirely consistent, but generally compatible (Matras 2002:254). As an example, (2) gives Sampson’s alphabet (1926:3):

(2) a â b ċ d ð e ě f g γ h i ĵ k k̇ χ l l̇ m n ŋ o p ṗ r ṙ s š t ṫ þ u v w ẇ y z ž

Sampson based his alphabet on that of Miklosich, who produced a 16 part dialectological survey of Romani in the late 19<sup>th</sup> century (Matras 2002:3), and whose system of writing the language, Sampson says ‘has generally been followed by modern Gypsy scholars’ (Sampson 1926:3). <š> and <ž> are the voiced and voiceless ‘open blade-point’ consonants (*ibid.* p.5), in modern terminology, postalveolar fricatives – IPA /ʃ, ʒ/. <č> and <ĵ> are postalveolar affricates (*ibid.* p.12), IPA /tʃ, dʒ/; the voiced consonant differs from Miklosich who used <dž> (*ibid.* p.3). <k̇, ṫ, ṗ> are the aspirated voiceless stops (*ibid.* pp.11,13,16), again differing from Miklosich’s <kh, th, ph> (*ibid.* p.3). Sampson also differs from Miklosich in using <χ> for his <ch>, the voiceless ‘open back’ consonant (velar fricative, IPA /x/); and <y> for <j> (IPA /j/). <â> represents the rounded vowel of English ‘not, naught’ (*ibid.*); <ě> the schwa (*ibid.* p.10); <γ> the voiced counterpart of <χ> and <ŋ> the velar nasal (*ibid.* p.12). Some of the letters represent sounds not in other dialects, which will therefore not be represented in the other orthographies examined below and are thus less relevant to the discussion: voiceless <l̇, ṙ, ẇ> and labiodental fricatives (only in English loanwords) <þ, ð> (*ibid.* p.3). The remainder of the letters are used roughly as in English.

Moving forward to more recent times, we can see that the majority of these conventions – particularly those of Miklosich, where he and Sampson differ – have remained. Matras (1999:488, 2002:254) lists the following main features of contemporary academic Romani transcriptions: use of <č, š, ž>, with the ‘wedge’ accent for postalveolars; Miklosich’s use of <h> to mark aspiration on voiceless stops (extended also to the voiceless postalveolar affricate <čh>, a sound not in the dialect studied by Sampson); and <x> for the velar fricative, differing from both Miklosich and Sampson (although <h> and <ch> are also used). In Matras’s own account of Romani phonology (Matras 2002:49-58) we can see some other of Miklosich’s graphemes in use – <dž> for the voiced postalveolar affricate and <j> for the semivowel. Matras also uses <c, dz> for the alveolar affricates, not in Sampson. In some areas conventions differ: palatalized consonants, also not present in Welsh Romani, may be represented in various ways – <t’>, <tj>, <ty> or <ć> for the voiceless stop; the uvular, as

opposed to trilled <r> which occurs in some dialects can be <ř> or <rr>; and schwa has numerous representations, of which the most common is <ə> (Matras 1999:488).

The academic conventions have influenced those of native speakers, both through the use by native speakers of dictionaries produced by linguists, and through the involvement of linguists in attempts at standardization, resulting in those attempts being ‘oriented’ towards the international linguistic conventions over those of national state languages (Matras 2002:254). It is to the orthographies used by native speakers that I now turn.

According to Matras (1999:482-3), ‘attempts to write Romani for purposes other than academic documentation go back at least to the first translation of parts of the Gospel into Romani in 1836’. Matras notes, however, that when the translation was finally published in 1911 it included a note in German at the front asking the reader to ‘distribute this book among the Gypsies, and to read it aloud to them’ – the Roma themselves were still presumed to be illiterate at the time. Hancock (1995:34), however, states that since ‘at least the last quarter’ of the 19<sup>th</sup> century, native speakers have ‘attempted’ to write their language, and cites the example of Russian Roma in the 1920s who wrote to American relatives using Cyrillic. Sampson (1926:viii) also refers to the scholar Francis Hindes Groome who corresponded with a Welsh Gypsy, John Roberts of Newtown, who could write Romani. The publication of Romani texts in the USSR was referred to above; Friedman (2005:163) calls this ‘the beginnings of native literacy’, but notes that ‘the experiment was short lived’. Writing by Roma for Roma remained scarce through the twentieth century, with ‘some, though few, political newsletters and private correspondence’ in the 1970s and onwards among ‘Romani political activists and intellectuals’; and an eventual ‘upsurge’ in publications after the fall of the USSR (Matras 1999:483).

With this increase in written Romani, attempts to standardize the orthographies in use began. Such attempts fall into two classes: those attempting to standardize the spelling of a particular dialect or regional group of dialects, and those attempting to create an orthography for all dialects. Supporters of creating one standard Romani language have mainly belonged to the ‘circle of a few dozen...regular participants in the framework of the International Roman Union formed in 1971’ (Matras 1999:487), and it is here that the idea of one alphabet to write all dialects has found most favour. Such an orthography was developed in the 1980s by Marcel Cortiade (also spelt Courthiade), and his proposal was adopted as the Union’s official alphabet at the 4<sup>th</sup> World Romani Congress in Warsaw, 1990, and called the ‘Romani common alphabet’ or ‘International Standard’ (Matras 1999:491, 2002:252; Friedman 1997:185; Kenrick 1996:118).

In many ways Cortiade’s alphabet resembles the academic conventions outlined above – using <h> for aspiration, <x> for the velar fricative, <j> for the semivowel and <c, dz> for the alveolar affricates (Friedman 1995:182-3; Hancock 1995:38-44; Hübschmannová & Neustupný 1996:101). He distinguishes between the unmarked <r> and the marked <rr>, uvular or long depending on dialect (Friedman 1997:188). Schwa is not given a representation, as it does not occur in all dialects (Friedman 1995:183).

The main features unique to Cortiade are as follows. Firstly, where academic transcriptions mark the postalveolars with a wedge accent: <č, čh, š, ž>, Cortiade uses the acute: <ć, čh, ś, ź> (Hancock 1995:44; Hübschmannová & Neustupný 1996:101). The exception to this is the voiced affricate, <dž>, which Cortiade writes as <ǰ> (Matras

2002:252). As this is a ‘polylectal’ alphabet (Matras 1999:491), intended for use by speakers of different dialects, <čh> and <ǰ> are in fact ‘archegraphemes’ (Matras 2002:252), intended to represent different pronunciations of the same original sound. This results in a problem when the sounds they represent have in fact merged with the fricatives <ś, ź> – the speaker will have no way of knowing which words originally had affricates and should therefore be spelt with <čh, ǰ>; they will ‘need to know the etymology of a word before deciding which symbol to write it with’ (Kenrick 1996:119).

The other unique feature of Cortiade’s orthography is the use of ‘morphophonemic’ symbols (Friedman 1997:186) or ‘morpho-graphs’ (Matras 1999:491) <θ> and <q> for use in case endings (also called postpositions). Case endings in Romani ‘show voice assimilation to the oblique endings of the noun to which they attach (dative *-ke/-ge*, locative *-te/-de*, ablative *-tar/-dar* etc.)’ (Matras 2002:79); the ‘morpho-graphs’ represent both the voiced and voiceless sounds, <θ> standing in for <t, d> and <q> for <k, g>. Thus *dadeske* ‘to the father’ and *dadenge* ‘to the fathers’ would be written *dadesqe* and *dadenqe*, and pronounced in various ways depending on dialect (Kenrick 1996:119). A further morpho-graph is <ç>, standing for <s, c> (Hancock 1995:44).

Cortiade’s system has been criticised for being inconsistent in its choice of which areas of dialect variation to reflect by introducing ‘archegraphemes’ (Matras 1999:491), and despite EU backing, has met little success (Matras 2002:252). The creation of a single orthography for all dialects is itself problematic, in the absence of a standard language, as Kenrick (1996:118) points out: readers educated in the state language will interpret the letters used for Romani as having the same value as when used for writing the state language, so for example the word *ja* would be read by a Swedish gypsy as /ja/ ‘yes’, by an English speaking gypsy as /dža/ ‘go’, and by a Castilian speaking gypsy as /xa/ ‘eat’.

This suggests that it would be easier for dialects in different areas to develop different orthographies based on the national language. This has happened, and been more successful than the ‘International Standard’. I will discuss a couple of the more well-documented examples

In Macedonia, standardization began with the publication of Jusuf and Kepeski’s *Romani gramatika*, ‘Romani Grammar’ in 1980, based a mixture of dialects (Friedman 1995:181, 1996:90). Yugoslavia at the time had two official scripts in use, Latin and Cyrillic; Jusuf and Kepeski used the Yugoslavian Latin alphabet as the basis for their orthography (Matras 1999:485-6), and this was maintained after Macedonian independence, despite Cyrillic becoming the national script of Macedonia. Friedman (1995) discusses the document produced at a conference for standardization organised by the University of Skopje and held in November 1992; (3) gives the alphabet as stated in that document (Friedman 1995:181):

(3) Aa Bb Cc Čč Čh/čh Dd Dž/dž Ee Ff Gg Hh Ii Jj Kk Kh/kh  
Ll Mm Nn Oo Pp Ph/ph Rr Ss Šš Tt Th/th Uu Vv Žž

We can see here numerous familiar symbols. The alphabet resembles the linguistic conventions discussed above in its use of <h> for aspiration, <j> for the semivowel and the wedge accent <ˇ> to mark postalveolars – the last of these also common in East European

orthographies (*ibid.* p.182). There is no marking of the velar fricative; Jusuf and Kepeski had proposed using <x>, but as the phoneme developed historically from /h/, and the distinction isn't present in all dialects (Friedman 1996:93-4), <h> stands for both phonemes here. The document also differs from Jusuf and Kepeski in omitting their proposed grapheme for schwa, <ä> (Friedman 1995:182). The alphabet has been relatively successful, and the 'basic principles...have remained consistent in almost all published literature' (Friedman 2005:166), with only a few areas of variation such as whether to represent palatalization of dentals and velars before front vowels (Friedman 1995:183, 1997:186, 2005:166), or whether to omit schwa as recommended by the standardization conference, or represent it with an apostrophe as in Macedonian orthography (Friedman 1996:92-3; 1997:185-6, 2005:166).

Another fairly successful orthography is that used in the Czech Republic and Slovakia. The *Svaz Cikánù-Romù* or Union of Gypsies-Roma, established 1969, had a Linguistic Commission which developed an orthography based on the spelling of Slovak and Czech. Table 1 gives the alphabet, with its Czech equivalent (Hübschmannová 1995:193,197; Hübschmannová & Neustupny 1996:100-1):

**Table 1: The Slovak-and-Czech Romani alphabet compared with Czech**

Romani	a	-	b	c	č	čh	d	d'	dz	dž	e	-	f	g	h	ch
Czech	a	á	b	c	č	-	d	d'	-	-	e	é	f	g	h	ch
Romani	i	-	j	k	kh	l	l'	m	n	ň	o	-	p	ph	-	r
Czech	i	í	j	k	-	l	-	m	n	ň	o	ó	p	-	q	r
Romani	-	s	š	t	t'	th	u	-	-	v	-	-	-	z	ž	
Czech	ř	s	š	t	t'	-	u	ú	ù	v	x	y	ý	z	ž	

Again, as in Macedonia, basing the alphabet on that of the state language automatically results in some graphemes being identical with those used by linguists – the wedge accent for postalveolars, <j> for the semivowel, <c> for the voiceless alveolar affricate. Noticeably different is the use of Czech <ch> rather than <x> for the velar fricative. On the other hand, the orthography does appear to draw on linguistic conventions in its use of <h> for aspiration. The graphemes <d', l', ñ, t'> represent palatalized consonants; here the orthography differs from Czech (Hübschmannová 1995:197).

There are many other similar examples of regional standardized orthographies: in Finland the Ministry of Education appointed an orthography committee in 1970, and the orthography produced became 'fairly established' (Granqvist 2006:54). In Austria in the 1990s linguists from the University of Graz worked with speakers of an endangered dialect, Roman, using questionnaires to work out the spelling preferences of the speakers and from this construct an orthography. The native speakers rejected diacritics proposed by the linguists and the resulting orthography was based on German (Matras 1999:486, 2002:253-4). More such examples are listed in Matras (2004:5-9).

In the absence of a standardized alphabet, speakers of Romani will generally adapt the spelling conventions familiar to them from the state language to writing Romani (Hancock 1995:34-5; Matras 2002:253). Hancock gives as an example a verse of a song, taken from

some printed lyrics accompanying a CD, using a German based orthography. (4) gives the first line, with the academic-style respelling Hancock provides:

- (4) a) Schej ben soste man chochawes  
b) Čhej phen soste man xoxaves

The linguistic spelling conventions can influence the spelling used, however – Matras (1999:488-9) describes the various degrees to which this can happen. The most commonly used academic convention is the marking of aspiration with <h>; this often the only addition to the spelling conventions of the majority language, but further adaptations may also occur; such compromises between national and academic conventions ‘testify to the international orientation’ of writers (Matras 2002:256).

In recent years the growth of the Internet has added another dimension to Romani orthography, allowing Roma from around the world to communicate, each using their own spellings but influencing one another (Matras 2002:257). The difficulty of writing letters with diacritics on a keyboard has led to the use of a largely English based orthography – such an orthography is exemplified in both Hancock (1995:43-4) and Lee (2005:5-11). Features common to Hancock and Lee which differ from academic conventions include use of <ts> for the voiceless alveolar aspirate <c>; <ch> for the postalveolar <č>; <sh, zh> for the fricatives <š, ž>; a following <y> for palatalization: <ny, ly>; and <y> for the semivowel <j>.

This concludes this brief survey of some of the many orthographies that have been used for Romani; there is much more that could be said.

## References

- Bakker, Peter & Hristo Kyuchukov (eds.). 2000. *What is the Romani language?* Hatfield: University of Hertfordshire Press.
- Friedman, Victor A. 1995. Romani standardization and status in the Republic of Macedonia. In Matras 1995, 177-188.
- Friedman, Victor A. 1996. Romani and the census in the Republic of Macedonia. *Journal of the Gypsy Lore Society Series 5*, 6. 89-101.
- Friedman, Victor A. 1997. Linguistic form and content in the Romani-language press of the Republic of Macedonia. In Yaron Matras, Peter Bakker & Hristo Kyuchukov (eds.), *The typology and dialectology of Romani*, 183-198. Amsterdam: John Benjamin's Publishing Company.
- Friedman, Victor A. 2005. The Romani language in Macedonia in the third millennium: Progress and problems. In Barbara Schrammel, Dieter W. Halwachs & Gerd Ambrosch (eds.), *General and applied Romani linguistics: Proceedings from the 6<sup>th</sup> International Conference on Romani Linguistics*, 163-173. Munich: Lincom Europa.
- Granqvist, Kimmo. 2006. (Un)wanted institutionalization: The case of Finnish Romani. *Romani Studies Series 5*, 16(1). 43-61.
- Hancock, Ian. 1995. *A handbook of Vlax Romani*. Columbus: Slavica Publishers.
- Hübschmannová, Milena. 1995. Trial and error in written Romani on the pages of Romani periodicals. In Matras 1995, 189-205.
- Hübschmannová, Milena & Jiří V. Neustupny. 1996. The Slovak-and-Czech dialect of Romani and its standardization. *International Journal of the Sociology of Language* 120. 85-109.
- Kenrick, Donald. 1996. Romani literacy at the crossroads. *International Journal of the Sociology of Language* 119. 109-123.
- Lee, Ronald. 2005. *Learn Romani*. Hatfield: University of Hertfordshire Press.
- Matras, Yaron (ed.). 1995. *Romani in contact: The history, structure and sociology of a language*. Amsterdam: John Benjamin's Publishing Company.
- Matras, Yaron. 1999. Writing Romani: The pragmatics of codification in a stateless language. *Applied Linguistics* 20(4). 481-502.  
<[http://applied.oxfordjournals.org/content/20/4/481.full.pdf + html](http://applied.oxfordjournals.org/content/20/4/481.full.pdf+html)> (5 March, 2011)
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2004. The future of Romani: Toward a policy of linguistic pluralism.  
<[http://romani.humanities.manchester.ac.uk/downloads/2/Matras\\_Pluralism.pdf](http://romani.humanities.manchester.ac.uk/downloads/2/Matras_Pluralism.pdf)> (5 March, 2011)
- Romani Wikipedia. <<http://rmy.wikipedia.org/wiki/>> (28 February, 2011)
- Sampson. 1926. *The dialect of the Gypsies of Wales*. Oxford: Oxford University Press.
- Sergievsy, M. V. & A. P. Barannikov. 1938. *Gypsy-Russian dictionary*. Moscow.  
<<http://liloru.ru/romanes/sergievsky1.htm>> (7 March, 2011)