Kimmo Granqvist

NOTES ON FINNISH ROMANI PHONOLOGY

0. INTRODUCTION

The aim of this paper is to describe some main features of the segmental phonology of the Finnish Romani language. This paper proposes a phoneme inventory of Finnish Romani. In addition to this, phoneme frequencies are calculated. Some phonotactic constraints are also described by using the methodology introduced by the Danish school of structural linguistics; position, vowel adherence, combination and combinability analyses are presented.

The material used in the present study consists of eight computerized corpora of totalling 52,332 words. The corpora are available at the Research Institute for the Languages of Finland. Five of the corpora used are SGML-coded dictionaries of Finnish Romani (Thesleff, Kronqvist, Jalkio, Temo and MNS). The dictionaries originate from quite different periods of time; the oldest one, the dictionary by Thesleff was published in 1911, while MNS was issued in 1971. The three text corpora (Bible1, Bible2, Bible3) include passages from the Bible, translated into Finnish Romani by various authors. The composition of the material used is shown in table (1).

Corpus	Overall size
Thesleff	7,577
Kronqvist	2,679
Jalkio	7,563
Temo	4,478
MNS	2,992
Bible1	5,086
Bible2	17,295
Bible3	4,662
Total	52,332

Table 1. The material used.

1. PHONEME SYSTEM

1.1. Vowels

Contemporary Finnish Romani has a eight-vowel-system, identical to the Finnish vowel system. The vowel system is presented in (1). Some examples of contrasts of the vowel phonemes are shown in (2):

(1)

	Front		Back	
	Unrounded	rounded	Unrounded	rounded
Close	i	ü		u
Mid-open	e	ø		o
Open	æ		a	

(2)

i — a	*lim* 'snot' — *lam* 'step'
ü — a	*süü* 'needle'— *saa* 'laugh'
u — a	*buur* 'lout'— *baar* 'fence'
ø — o	*øm* 'sensitive'— *om* 'if'
æ — a	*bæær* 'field' — *baar* 'fence'
o — i	*raklo* 'non-gypsy boy' — *rakli* 'non-gypsy girl'

Note that some of the contrasts, in particular æ — a, ø — o, are weak.

In addition to the basic five vowels /a, e, i, o, u/ common to all dialects of Romani, the front vowels /ü, ø, æ/ were adopted into the language, presumably due to Hungarian and Scandinavian influency (Valtonen 1968). The distribution of /ü, ø, æ/ is limited into a group of chiefly germanic and scandinavian loan items, such as those listed in (3):

(3) *büükæ* 'laundry' < Germ., Sw. *byk*
bæri 'hill' < Sw. *berg*
hüøg 'high' < Sw. *hög*
lüørdæ 'Saturday' < Scand. *lørg*, Sw. *lördag*
nüødæ 'need' < Scand. *nød*, Sw. *nöd*
stükkøs 'piece' < Germ. *stykke*, Sw. *stycke* etc.

The etymologies are presented according to Valtonen (1972). In the older layers of the Finnish Romani vocabulary, /ü, ø, æ/ are rare; the examples shown in (4) are presented based on Valtonen (1968).

(4) *tʃæi* 'girl'
dæi'mother'
gææji 'non-gipsy woman'
phüüli 'widow'
ræi 'lord'
t^{h}üøøli 'cigarrette'

It must be pointed out that also the parallel form of these words *tʃaj* 'girl', *daj* 'mother', *gaaji* 'non-gypsy woman' and *raj* 'lord' occur.

1.2. Consonants

Finnish Romani has 23 consonant phonemes. The consonant system is presented in (5) and some examples of the contrasts of the consonants phonemes in (6).

(5)

	Lab.	**Dent.**	**Alv.**	**Vel.**	**Lar.**
Stops					
Voiceless unaspirated	p	t		k	
Voiceless aspirated	p^h	t^h		k^h	
Voiced unaspirated	b	d		g	
Fricatives					
Voiceless spirants		f	–	x	h
Voiced spirants	v	–		–	–
Voiceless sibilant			s		
Voiced sibilant		z			
Voiceless sibilant				ʃ	
Voiced sibilant			–		
Affricates					
Voiceless			tʃ		
Voiced			dʒ		
Nasals	m	n			
Lateral		l			
Tremulant		r			
Semivowel		j			

(6)

p — r	*sap* 'snake' — *sar* 'so'
p — t	*passavaa* 'suit' — *tassavaa* 'warm up'
k — m	*tʃerko* 'strong' — *tʃermo* 'worm'
p^h — p	*p^hariba* 'flounces' — *pariba* 'changing'
t^h — t	*t^huu* 'smoke' — *tu* 'thou'
k^h — k	*k^haan* 'smell' — *kaan* 'now'
b — tʃ	*baar* 'fence' — *tʃaar* 'grass'
d — b	*doori* 'there' — *boori* 'daughter-in-law'
g — k	*goodi* 'brains' — *koodi* (pl. < kooda) 'resin'
f — p	*finnos* 'Finn' — *pinnos* 'nail'
x — h	*xastiba* 'poverty' — *hastiba* 'hurry'
h — t	*haagiba* 'wish' — *taagiba* 'paralysis'
h — v	*hast* 'at once' — *vast* 'hand'
s — z	*sar* 'so' — *zar* 'hair'
ʃ — s	*beʃto* 'embroidered' — *besto* 'best'
x — m	*hool* 'deep' — *mool* 'wine'
n — m	*naal* 'in front of' — *maal* 'friend'
l — r	*hamlos* 'corn' — *hamros* 'hammer'
r — d	*duur* 'far away' — *duud* 'pupil'
j — v	*jaaro* 'egg' — *vaaro* 'flour'

The phoneme /n/ has two allophones [n, ŋ]. Due to nasal assimilation, the allophone [ŋ] is found before the velars /k, g/.

The contrast between the phonemes /ʃ/ and /x/ is has minimal/no functional load. Typically, /ʃ/ and /x/ alternate. In Eastern dialects, /ʃ/ is substituted for /x/, e.g. /xeel/ → /ʃeel/ 'hundred', /xeero/ → /ʃeero/ 'head', /xieelalo/ → /ʃieelalo/ 'cold', /xlieva/ → /ʃlieva/ 'ladle' etc. Additionally, in the Eastern dialects the /dj/, or /j/ occurs instead of the affricate /dʒ/, e.g. /dʒeeno/ → /(d)jeeno/ 'man', /dʒuuli/ → /(d)juuli/ 'woman' etc.

Influenced by the Finnish phonology, Finnish Romani has been subject to other phonological changes, too, which are specific either to certain registers, or to certain idiolects:

(i) As in spoken Finnish, the voiced stops /b,d,g/ often are realized as /p,t,k/, e.g. /iego/ → [ieko] 'own' (Hedman 1996: 27-8)

(ii) Since Finnish lacks aspirated stops, the unaspirated voiceless stops /p, t, k/ tend to substituted for the aspirated voiceless stops /p^h, t^h, k^h/, e.g. /p^hurano/ → [purano] 'old' (Hedman 1996: 32)

(iii) The alveolar sibilant /ʃ/ and the affricate /tʃ/ tend be replaced by /s/ or /ts/, as Finnish lacks /ʃ, tʃ/. /ts/, or /s/, is frequently substituted for the voiced affricate /dʒ/, which Finnish does not utilize. E.g. /dʒuklo/ → [suklo] 'dog' (Hedman 1996: 36).

(iv) The voiceless velar spirant /x/ is often replaced by /h/. E.g. /xatʃiba/ → [hatsiba] 'burning' "cigarette".

1.3. Distinctive Features

Figure 1. *Vowels*

	i	ü	u	e	ø	o	æ	a
hi	+	+	+	-	-	-	-	-
low	-	-	-	-	-	-	+	+
bk	-	-	+	-	-	+	-	+
rnd	-	+	+	-	+	+	-	-

Figure 2. *Consonants*

	p	t	k	p^h	t^h	k^h	b	d	g	f	x	h	v	s	z	ʃ	tʃ	dʒ	m	n	l	r	j
cons	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
son	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+
int	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-
nas	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
lat	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
vd	-	-	-	-	-	-	+	+	+	-	-	-	+	-	+	-	-	+	+	+	+	+	+
hi	-	-	+	-	-	+	-	-	+	-	+	-	-	-	-	+	+	+	-	-	-	-	-
cor	-	+	-	-	+	-	-	+	-	-	-	-	-	+	+	+	+	+	-	+	+	+	+
str	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	+	+	+	-	-	-	-	-
HSP	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

2. PHONEME FREQUENCIES

The phoneme frequencies are compiled on the basis of all eight corpora used in this study. As normally in Finnish phonology, the long vowel and consonants have been considered two separate segments. Thus, for instance /kaalo/ 'dark' has been analyzed as consisting of the five phonemes /k + a + a + l + o/ instead of four: /k + a: + l + o/.

Most phonemes are corresponded by one and only one grapheme in the corpora, which makes the calculation of the phoneme frequencies quite straightforward. Character conversions are only required in the cases shown in table (2).

Grapheme	Phonemic representation
å, o	o
ü, y	ü
ph	p^h
th	t^h
kh	k^h
ĥ, ch	x
š, sch, sh	ʃ
č, tsh, tsch	tʃ
ž, dž, dsch, dš	dʒ

Table 2. Grapheme-phoneme correspondences

2.1. Vowels

Table (3) clearly shows the primacy of five basic vowels /a, e, i, o, u/ compared to the loan vowels /ü, ø, æ/. The five basic vowels cover together 97.97 % of all the vowel occurrences in the corpora.

Vowel	N	%
a	58,276	35.34
e	30,179	18.30
o	29,918	18.03
i	29,139	17.67
u	13,929	8.45
ø	1,318	0.80
ü	1,306	0.79
æ	842	0.51
Total	164,907	99.89

Table 3. Phoneme frequencies of Finnish Romani vowels

2.2. Consonants

Twelve of the Finnish Romani 23 consonants are alveodentals. Phonologically I treat them are as dentals, for simplicity reasons. Expectedly, the frequency of the alveodentals is very high in the corpora, 59.26% of all consonants. The proportion of the labials is 22.01% and the proportion of the palatovelars 16.69%. The only laryngeal /h/ has a corpus frequency of no more than 1.77% (Table 4).

The examination of the phoneme frequencies also shows tendencies related with phonological properties other than the place of articulation. As expected, generally the more unmarked phoneme occurs more frequently than its more marked counterpart. Thus:

(i) The (unmarked) voiceless obstruents are more common than the corresponding (marked) voiced obstruents, with exception, however, of /b, v/, which are found more often in the corpora than /p, f/.
(ii) Both unaspirated voiceless and voiceless stops are used clearly frequently than the voiceless aspirated stops.
(iii) Dental stops are more commonly used than the corresponding affricates.
(iv) The liquid /r/ ([-lat]) occurs more often than the /l/ ([+lat]).

Consonant	N	%
s	22,453	12.40
n	21,281	11.75
k	19,374	10.70
r	16,342	9.02
t	14,258	7.87
l	13,796	7.62
b	12,191	6.73
v	9,946	5.49
d	9,427	5.21
m	9,134	5.04
g	5,968	3.29
p	5,224	2.88
x	4,905	2.71
j	3,435	1.90
h	3,197	1.77
tʃ	3,146	1.74
p^h	1,701	0.94
f	1,665	0.91
dʒ	1,543	0.85
ʃ	875	0.48
k^h	475	0.26
z	385	0.21
t^h	372	0.21
Total	181,093	99.99

Table 4. Phoneme frequencies of Finnish Romani consonants

3. PHONOTACTICS

3.1. Position Analysis

Fischer-Jørgensen (1952) suggests methods for the classification of phonemes, based on different positions. She classifies the phonemes according to their tendency to occur in syllable initial and final positions, or in the position next to the vowel in CC clusters etc. The position analysis has applied in many studies of Danish and Swedish phonology, for instance.

In the present study I have defined different positions in both initial and final CC and CCC clusters, depending on the distance of the consonant from the vowel following or preceding the cluster. The examples in (7) show how the positions are assigned in initial and final CCC clusters.

(7)

#	3 s	2 k	1 r	V
V	1 s	3 t	3 r	#

The positions are denoted by arabic numbers, whereas the length of the cluster is expressed by roman number, e.g. 3III, 2III, 1III, 2II, 1II. Thus, 1III and 1II are the classes of phonemes which can occur immediately adjoining the vowel, whereas 3III and 3II are the classes next to word boundary.

In figures (3–6) the phonemes are ranked according to their tendency to occur in different positions. In initial CC clusters the sibilant /s/ and the stops /p, d, t, k, b, g/ are most frequent in position 2II. On the other hand three groups of sound are very unlikely to be found in this position:

(i) the aspirated stops, which generally do not occur in many consonant clusters;
(ii) the fricatives /x, ʃ/;
(iii) the nasals /m, n/.

In position 1II, that is, next to the vowel, the sonorants /r, l, j, n/ are most frequent.

The results describing the final CC clusters are much harder to interpret. Here also the sonorants are favoured in position 1II next to the vowels, while the fricatives /x, ʃ/ are least common in this position.

As for initial CCC clusters, the results resemble those found in many studies of other IE languages (e. g. Sigurd 1965; Setatos 1971). The position analysis shows that the sibilant /s/ is dominant.

Figure 3. *Initial CC*

2 II		1 II		V
p^h	1	tʃ	1	
t^h	1	b	1	
m	1	f	1	
x	2	g	1	
ʃ	2	k	1	
h	2	m	1	
n	2	p	1	
v	2	z	1	
f	3	t	2	
g	4	h	3	
b	5	v	3	
k	5	s	4	
t	5	x	5	
d	7	n	5	
p	7	j	8	
s	8	l	9	
		r	10	

Figure 4. *Final CC*

V	1 II		2 II	
	x	1	b	1
	ʃ	1	d	1
	g	1	g	1
	h	1	p	1
	j	1	r	1
	t	1	z	1
	d	1	ʃ	2
	l	2	m	2
	m	2	x	3
	s	3	k	3
	k	3	s	3
	r	6	j	3
	n	10	h	4
			t	8

Figure 5. *Initial CCC*

3 III		2 III		1 III		V
d	1	k	1	l	1	
p	2	p	1	j	1	
s	3	r	1	r	4	
		t	1			
		v	1			
		z	1			

Figure 6. *Final CCC*

V	1 III		2 III		3 III	
	l	1	dʒ	1	x	1
	n	5	s	1	ʃ	1
			d	2	k	1
			g	2	s	1
					h	2

The first segment (1III) of final CCC clusters can only be either the lateral /l/ or, much more commonly (f=5), the nasal /n/. In the two other positions only obstruents are found. In position 2III the voiced stops /d, g/ are slightly favoured, while in position 3III /h/ is a bit more common than the sibilants /s, ʃ/ and the velar /k, x/.

3.2. Vowel Adherence

The term "vowel adherence" refers to the tendency of the consonants to occur close to the vowel. The tendency is assumed to be equal to the difference between its frequency in position 1II and 2II. The following formula has been used to calculate the vowel adherence:

(8) Ai = f1IIi - f2IIi,

in which Ai is the tendency of the phoneme i to occur close to the vowel, f1IIi the frequency of the phoneme i is position 1II, and f2IIi the frequency of the phoneme i in position 2II. The frequencies used here are not real corpus frequencies. Instead we make use of the number of clusters, in which the phoneme i is found in different positions. For CCC clusters, following the suggestion by Sigurd (1965: 50) the positions 2III and 3II are treated as one position.

Figure (7) shows that in initial CC clusters, the stops /d, p, k, b/ (that are most consonantal as segments) and the sibilant /s/ have the weakest tendency to occur adjoining the vowel. The sonorants /r, l, j, n/ – which are least consonantal – have on the other hand the strongest tendency to occur close to the vowel. Generally speaking, voiced fricatives have higher vowel adherence values than voiceless fricatives and stops. As figures (8–10) indicate, the results are quite similar for other types of clusters examined, too.

Figure 7. *Initial CC*

			tʃ								
			v					k			
			x	h		p^h	f	t	b		
r	l	j	n	z	m	t^h	ʃ	g	s	p	d
10	9	8	3	1	0	-1	-2	-3	-4	-6	-7

Figure 8. *Final CC*

				k	p			
				d	b			
				g	z	j		
n	r	l	m	s	ʃ	x	h	t
10	5	2	1	0	-1	-2	-3	-7

Figure 9. *Initial CCC*

	t			
		k		
		d		
	l	v		
j	r	z	p	s
2	1	-1	-2	-4

Figure 10. *Final CCC*

	k		
		ʃ	
		dʒ	d
		h	g
n	l	x	s
5	1	1	-2

3.3. Combination Analysis

A fairly different approach has been proposed by Spang-Hansen (1953), who is interested only in what kind of consonant clusters are found in language, and what kind of clusters do not occur. Charts that indicate which phonemes can combine, have been used for many studies of phoneme distribution (such as Pike 1947; Aurén 1869; Lyttkens/Wulff 1885; Porru 1939; Bjerrum 1944). In figures (11–13), which only represent two-consonant-clusters, the consonants are placed so that the first member of the cluster is found in the left column.

3.3.1. Initial Consonant Clusters

In all 54 different initial two-consonant-clusters are found in the corpora in 5,966 lexical items. 60% of these CC clusters belong to the type:

(9)

$$\begin{bmatrix} +\text{con} \\ -\text{son} \end{bmatrix} \begin{bmatrix} +\text{con} \\ +\text{son} \\ -\text{nas} \end{bmatrix}$$

Stop + liquid clusters (11 / 45.65%) are clearly more common in initial position than fricative + liquid clusters (7 / 10.92%). The cluster type obstruent + nasal is rare: the overall frequency of the three stop + nasal clusters /pn, kn, gn/ found in initial position is no more than 1.70%. Fricative + nasal clusters /sm,sn/ are a bit more common (2.95%).

Two heterogeneous ([αint -αint]) types of clusters made by two obstruents are also relatively common:

(i) stop + fricative (15 / 10.36%)
(ii) fricative + stop (5 / 22.20%)

The following are the most important phonotactic restrictions that apply in initial two-consonant-clusters:

(10)

$$* \begin{bmatrix} +\text{con} \\ -\text{son} \\ \alpha\text{int} \end{bmatrix} \begin{bmatrix} +\text{con} \\ -\text{son} \\ \alpha\text{int} \end{bmatrix}$$

'Clusters of two stops or two fricative are not allowed in initial position.' The only exception this rule is made by the cluster /sv/, which is quite rare in the corpora (4.17%).

Figure 11. *Initial CC*

	p	t	k	pʰ	tʰ	kʰ	b	d	g	f	x	h	v	s	z	ʃ	tʃ	dʒ	m	n	l	r	j
p										pf	px			ps						pn	pl	pr	pj
t											tx			ts								tr	tj
k											kx		kv							kn	kl	kr	
pʰ																						pʰr	
tʰ													tʰv										
kʰ																							
b											bx	bh									bl	br	bj
d											dx	dh		ds	dz							dr	dj
g												gh								gn	gl	gr	
f																					fl	fr	fj
x																					xl		xj
h																					hl		hj
v																						vr	vj
s	sp	st	sk				sb						sv						sm	sn	sl		
z																							
ʃ		ʃt																			ʃl		
tʃ																							
dʒ																							
m																				mn			
n									ng								ntʃ						
l																							
r																							
j																							

Figure 12. *Medial CC*

	p	t	k	p^h	t^h	k^h	b	d	g	f	x	h	v	s	z	ʃ	tʃ	dʒ	m	n	l	r	j
p	pp	pt											pv	ps						pn	pl	pr	pj
t	tp	tt	tk						tg	tf	tx		tv	ts					tm	tn	tl	tr	tj
k	kp	kt	kk		kt^h		kb	kd	kg		kx		kv	ks				kdʒ		kn	kl	kr	kj
p^h																						phr	
t^h																							
k^h																							k^hj
b							bb						bv							bn	bl	br	bj
d								dd					dv	ds	dz					dn	dl	dr	dj
g		gt					gb	gd	gg				gv							gn	gl	gr	gj
f		ft								ff										fn	fl	fr	fj
x	xp	xt	xk	xp^h				xd			xx		xv			xʃ	xtʃ			xn	xl		xj
h		ht	hk					hd	hg			hh	hv							hn	hl	hr	hj
v													vv								vl	vr	vj
s	sp	st	sk	sp^h			sb	sd	sg		sx		sv	ss			stʃ	sdʒ	sm	sn	sl		sj
z							zb	zd												zn			zj
ʃ		ʃt	ʃk										ʃv			ʃʃ					ʃl	ʃr	ʃj
tʃ																							tʃj
dʒ													dʒv										dʒj
m	mp	mt	mk				mb	md					mv	ms					mm	mn	ml	mr	mj
n	np	nt	nk				nb	nd	ng	nf	nx	nh	nv	ns	nz	nʃ	ntʃ	ndʒ	nm	nn	nl	nr	nj
l	lp	lt	lk	lp^h			lb	ld	lg		lx	lh	lv	ls		lʃ	ltʃ	ldʒ	lm	ln	ll	lr	lj
r	rp	rt	rk			rk^h	rb	rd	rg		rx	rh	rv	rs	rz	rʃ	rtʃ	rdʒ	rm	rn	rl	rr	rj
j		jt	jk				jb	jd												jn	jl	jr	jj

Figure 13. *Final CC*

	p	t	k	p^h	t^h	k^h	b	d	g	f	x	h	v	s	z	ʃ	tʃ	dʒ	m	n	l	r	j
p																							
t																							
k		kt									kx												
p^h																						phr	
t^h																							
k^h																							
b																							
d															dz								
g												gh											
f																					fl	fr	fj
x		xt																			xl		xj
h		ht																			hl		hj
v																						vr	vj
s		st	sk																			sr	
z																							
ʃ		ʃt																			ʃl		
tʃ																							
dʒ																							
m							mb													mn			mj
n		nt	nk					nd	ng		nx	nh		ns		nʃ		nm					nj
l		lt																					lj
r	rp	pt	rk								rx			rs		rʃ							
j						js																	

(11)

$$\begin{bmatrix} +\text{con} \\ -\text{son} \\ \alpha\text{vd} \end{bmatrix} \begin{bmatrix} +\text{con} \\ -\text{son} \\ \alpha\text{vd} \end{bmatrix}$$

'In initial two-obstruent-clusters both obstruents must be either voiceless or voiced.'

(12)

* [+son] /#_C

'Sonorants cannot occur in position 2II of initial CC clusters.'

(13)

$$* \begin{bmatrix} +\text{con} \\ -\text{son} \\ +\text{int} \\ +\text{str} \end{bmatrix} \; / \{\#\_C, \; \#C\_\}$$

'Affricates cannot occur in initial CC clusters'

(14)

* [+HSP] / {#_C, #C_}

'Aspirated stops cannot occur in initial CC clusters'

3.3.2. Medial Consonant Clusters

As expected, a clear majority of the CC clusters found in the corpora are medial. In all there 201 different medial CC clusters. In the corpora the total number of occurrences of medial CC clusters is 28,156.

Two main types of medial CC clusters can be distinguished:
(i) long consonants
(ii) C + stop clusters

3.3.2.1. Long Consonants

Seventeen of the 201 clusters are made up by long consonants. Long consonants are very frequent in the corpora, representing about one third (35.51 %, N=8,418) of the medial data. All consonants except for the aspirated stops and the affricates can be lengthened. The probability of different consonants to be lengthened varies, however, considerably, as shown in table (5):

Sequence	N	% of medial CC
kk	1,938	6.89
nn	1,430	5.08
ss	798	2.83
tt	790	2.81
....		
ff	76	0.27
ʃʃ	41	0.15
vv	8	0.03
jj	1	0.004 ?

Table 5. Consonant lengthening in the corpora

As for their corpus frequency, the long consonants constitute a majority of medial CC clusters, in which both members have the same manner of articulation (table 6).

Sequence	Long consonants	%
stop + stop	3,710	98.70
fric. + fric.	1,114	69.06
nasal + nasal	2,018	91.51
liq. + liq.	1,175	98.66
semiw. + semiw.	1	100.00

Table 6. Long consonants and two-consonant clusters with both member sharing the manner of articulation

The clusters /mn/ (N=182) and /rl/ (N=15) are quite frequently used in the corpora. Otherwise, in most clusters of two different consonants with the same manner of articulation there is a morpheme boundary inbetween. E. g. /moolik#boskero/ 'miller', /buut#gongi/ 'many times', /jeek#gong/ 'once', /uutan#maan/ 'without me', /hin#maxkar/ 'is in the middle'. Note, however: /aptieka/ 'chemist's shop', /bulriba/ (< bultriba) 'knock' etc.

3.3.2.2 C + stop clusters

Another commonly found medial CC cluster type is made by consonant + stop clusters. Most of these clusters consist of a fricative and a stop: e.g. /ft, xp, xt, xk, ht, hk, sp, st, sk, sb, sd, sg/. Nasal + stop and liquid + stop clusters are clearly more rare in corpora. E.g. /mp, mt, mk, mb, md, np, nt, nk, nb, nd, ng; lp, lt, lk, rp, rt, lk/. Except for /kt^{h}/, /k^{h}j/, aspirated stops do not occur in medial CC clusters. Affricates only can occur as second member in medial two-consonant-clusters.

3.3.3. Final consonant clusters

In final position, 31 different two-consonant clusters are found in the corpora. The number of the lexical items containing final CC clusters is far more restricted than the number of the items with initial CC clusters. The data of final two-consonant-clusters are small, only 446 words.

Finnish Romani handles final CC clusters in a way somewhat different from initial CC clusters. In the corpora 99.11% of final CC clusters represent the type: C + obstruent.

As table (7) shows, the stops are more common than fricatives in position 2II. The most frequently found type of C + obstruent clusters is nasal + stop (5 / 45.29%), but fricative + stop clusters are quite common, too (5 / 28.03). Here also the clusters of two stops or two fricatives are very rare; thus the same limitation *[+con, -son, αint] [+con, - son, αint] we pointed out above for initial CC applies here, too.

Other types of final CC clusters are marginal. The corpus frequency of the final sonorant /j/ clustert /lj/ is only 0.67% (N=3) of all occurrences of final CC clusters. The only final nasal + nasal cluster /nm/ was found once (0.22%). Basically we can say that sonorants do not occur as last segments of final two-consonants-clusters. They can only be found adjoining the vowel.

Sequence	N. of seq.	N of seq.	N in corpora %
stop + stop	1	1	0.22
stop + fricative	3	3	0.67
fricative + stop	5	125	28.03
fricative + fricative	1	1	0.22
nasal + stop	5	202	45.29
nasal + fricative	4	26	5.83
liquid + stop	4	56	12.56
liquid + fricative	3	25	5.61
semiw. + fricative	1	1	0.22

Table 7. Final C + obstruent clusters

3.4 Combinability

The term "combinability" refers to the ability of the consonant to combine with other consonants. Combinability is equivalent to the number of consonant clusters, in which the consonant occur. To calculate the combinability of different Romani consonants, I make use of the following formula:

(15)
$$C_i = \sum_{J=1}^{n} f_j n_i$$

in which C_i is the combinability of the phoneme i. f_jn_i is the frequency of the phoneme i in position $_j$n, and n is the number of the segments in the cluster. Figures (14–17) show results calculated on the basis of initial and final CC and CCC data. The figures indicate that in both initial CC and CCC the sibilant /s/ and tremulant /r/ are the most combinable consonants. For final CC and CCC, the results are different: the nasal /n/ obtains here the highest combinability.

Figure 14. *Initial CC*

s: 12
r: 10
j: 9
l: 9
p: 8
t: 7
d: 7
x: 7
n: 7
k: 6
b: 6
g: 5
v: 5
f: 4
ʃ: 2
tʃ: 2
m: 2
p^h: 1
t^h: 1

Figure 15. *Final CC*

n: 10
t: 9
r: 8
k: 7
s: 6
h: 5
j: 5
x: 4
m: 4
l: 4
d: 3
ʃ: 3
p: 2
g: 2
b: 1
z: 1

Figure 16. *Initial CCC*

s: 4
r: 4
p: 2
j: 2
t: 1
k: 1
d: 1
v: 1
z: 1
l: 1

Figure 17. *Final CCC*

n: 5
s: 3
h: 3
t: 2
d: 2
g: 2
l: 2
k: 1
b: 1
x: 1
ʃ: 1
dʒ: 1

REFERENCES

Aurén, J.A.	1896	*Bidrag till Svenska Språkets Ljudlära*, Lindköping.
Bjerrum, A.	1944	*Fjoldemålets lydsystem*, Copenhagen.
Fischer-Jørgensen, E.	1952	On the Definition of Phoneme Categories on Distributional Basis, *Acta Linguistica* 7: 8–39.
Hedman, H.	1996	*Romanikielen kielioppiopas. Sar me sikjavaa romanes*, Jyväskylä.
Lyttkens, I.A./Wulf, F.A.	1885	*Svenska Språkets Ljudlära och Beteckningslära jämte en afhandling om Aksent*, Lund.
Pike, K.L.	1947	*Phonemics*; Ann Arbor.
Porru, G.	1939	Anmerkungen über die Phonologie des Italienischen, *TCLP* 8: 187–208.
Setatos, M.	1971	*Fonologia tis koinis neollinikis*, Athens.
Sigurd, B.	1965	*Phonotactic Structures in Swedish*, Ph.D. Thesis.
Spang-Hansen, H.	1953	Analyse ved hjælp af kombinationsdiagrammer, *Selskab för Nordisk Filologi, Aarsberetning* 1951–52: 22–26.
Valtonen, P.	1968	*Suomen mustalaiskielen kehitys eri aikoina tehtyjen muistiinpanojen valossa*, Licentiate Thesis.
	1972	*Suomen mustalaiskielen etymologinen sanakirja*, (Tietolipas 69), Helsinki.

Kimmo Granqvist
Research Institute for the Languages of Finland
Helsinki