INTRASENTENTIAL CODESWITCHING
IN THE SPEECH OF FINNISH ROMA. A CASE STUDY

KIMMO GRANQVIST
The Research Institute for the Languages of Finland

1. INTRODUCTION

This paper is a case study of intrasentential codeswitching (henceforth CS) in the speech of Finnish Roma. I use here the term codeswitching as a cover term for codeswitching and codemixing that some scholars make a distinction between. The objective of this paper is to provide a provisional comparative analysis of Romani-to-Finnish and Finnish-to-Romani CS. In particular, I shall consider:

        a) the functions of CS,
        b) the frequency and the length of the switches,
        c) the lexical categories of the single-lexeme-switches, and
        d) lexical items consisting of both Finnish and Romani morphemes

Additionally, some discussion will be devoted to borrowing, which is closely related to CS.

The theoretical framework chosen for this paper is the Matrix Language Frame Model (henceforth MLF) proposed by Myers-Scotton (1993b). In the MLF model, the language that sets out the grammatical frame is the Matrix Language (henceforth ML), while the other language participating in CS is called Embedded Language (henceforth EL). The grammatical frame is defined as morpheme order and system morphemes (inflections,

system words). The system morphemes contrast with the content morphemes (nouns, adjectives, verbs).

The ML frames a projection of the complementizer (henceforth CP). A CP is a clause with a complementizer (COMP); null elements are allowed in the CP, including the complementizer.

A CP may contain three types of constituents:

a) Mixed ML+EL constituents, in which all EL morphemes are congruent, i.e. their grammar does not clash with the ML grammar

b) EL islands, which only consist of EL morphemes framed by the EL grammar

c) ML islands, which only consist of ML morphemes framed by the ML grammar.

Intrasentential CS is defined as at least one constituent with EL morphemes co-occurring with at least one constituent with ML morphemes within the same CP. In the other words, intrasentential CS takes place within CPs that contain either at least one mixed ML+EL constituent, or at least one ML island and at least one EL island.

The material used in this study is based on three hours of recorded speech in Finnish Romani/Finnish. The tape-recordings were carried out in 1995 by Juhani Pallonen (the Research Institute for the Languages of Finland) during an old Roma language seminar. On the tapes, Miranda Vuolasranta and Tuula Åkerlund, both of them Roma themselves, three elderly Roma women, born in 1920, 1925, and 1928. The interviews were transcribed from the tapes partially by myself and partially by Hellevi Hedman-Valentin. The transcription was quite broad, as the aim was to provide material mainly for lexical and syntactic studies. The size of the resulting corpus is 20,111 words.

Both the interviewers and the interviewees represent the Karelian sub-dialect of Finnish Romani characterised, for instance, by (i) the tendency to retain [ʃ], e.g. [ʃeːl] 'hundred', (ii) the occurrence of [dj], [j] instead of the standard Finnish Romani affricate [dʒ], e.g. [djuːli], [juːli] pro [dʒuːli] 'woman', (iii) the fronting of [a] to [æ], eg. [tʃæi] pro [tʃai] 'non-Roma girl', and some lexical features such as the use of [siːvo] instead of [gaːjo] 'non-Roma man' etc. (Valtonen 1968: 246-250; Koivisto 1987; Hedman 1996).

The interviewers are teachers of Romani, with good skills in Romani. Both interviewers used, however, a learned and quite formal code, which is uninteresting from the CS point of view. The three interviewees used a more

casual code and provided a good bilingual material that shows much of the CS. While all three interviewees have a relatively good proficiency in Finnish Romani, there are several indications in the material that Finnish is the better known language for them (perhaps dormant bilingualism). The question of language proficiency is essential, because in Finland, the use of Romani has been degrading quite rapidly for a long time. Kopsa-Schön (1996: 44) states that only the elderly Roma (ca. over 65 years of age) were able to communicate fluently in Romani. The middle-aged (32-64 years) Roma still showed satisfactory skills, but the youngsters (under 31 years) she interviewed no longer knew the language.

In the speech of the three elderly speakers, 739 mixed CPs were found (641 with Romani as the ML, and 98 with Finnish as the ML), which I chose for a closer examination. The ML of the mixed CPs was identified using methods suggested by Myers-Scotton (1997): as many of the syntactic properties of Romani, such as the word order, are influenced by Finnish, the emphasis was laid on the language of the system morphemes (The System Morpheme Principle). Occasionally, this was problematic as some of the system morphemes are Finnish loan items. When necessary, other methods were also used to identify the ML, such as counting morphemes, as the ML is usually the source of more morphemes.

2. CS WITH ROMANI AS THE MATRIX LANGUAGE

The use of Finnish Romani has been a for long time limited to only a few situations (at home or as a secret language) (Valtonen 1968: 242-3). The lexicon has remained very small. The largest dictionary of Finnish Romani published so far (Thesleff 1901) contains ca. 7,500 entries, of which over 5,000 are periphrases. Koivisto's Romani-Finnish-English dictionary (1994) has about 5,500 entries, 1,800 of them periphrases. At the time of my writing this, Koivisto is preparing a larger Finnish-Romani dictionary, which will contain about 25,000-30,000 lexical entries. The effects of attrition have long been visible in the speech of the Finnish Roma (cf. Valtonen 1968; on attrition, see Myers-Scotton 1997: 225). Many speakers seem to have lost the consistency of lexeme and grammatical frames. Clearly, one of the functions of the CS from Romani to Finnish is to help the speakers to

compensate for their deficiency not being able to express themselves in Romani.

In the corpus used for this study, there are far more switches from Romani to Finnish than vice versa. Romani is the ML in 87 % of the mixed CPs. This kind of tendency for unidirectionality is characteristic for CS.

CP length	Frequency	%
1	467	72.07
2	109	16.82
3	41	6.33
4	15	2.31
5	4	0.62
6	2	0.31
7	2	0.31
8	1	0.15

Table 1. Length and frequency of switches from Romani to Finnish.

Table (1) shows expectedly that a vast majority of the intrasentential switches found are single-occurring Finnish lexemes. Stretches longer than two words are very rare. Thus, the guidelines are quite similar to those provided by Cook (1991), who outlines that (intrasentential and intersentential) CS consists of 84 % single word switches, 10 % phrase switches and 6 % clause switches.

Category	Frequency	%
Adjectives	21	4.50
Particles, adverbs	247	52.89
Nouns	110	23.55
Numerals	3	0.64
Pronouns	42	8.99
Verbs	44	9.42
Total	467	100.00

Table 2. Single occurring Finnish lexemes classified according to their category.

Table (2) classifies the single occurring EL lexemes according to their grammatical category. More than half of the single-lexeme-switches are different kinds of particles/adverbs, such as those in (1) In the examples, bold print is used for Finnish. As a category, the adverbs have been in general most open to Finnish interference. Many of the Finnish conjunctions have been fully integrated into Romani, so that they can be used along with or instead of the original Romani ones (Valtonen 1968: 172). These are, however, to be considered borrowings, rather than examples of CS. They will be discussed more thoroughly below.

(1)   nii joon **heti** tenkana
         ('so they think immediately')

         niin isä oli, mango daat hin **ihan** finitiko
         ('so father was, my father is totally Finnish')

         doori naa **muutenkaan** tšiňko
         ('it wasn't good even otherwise there')

Other categories are generally less susceptible to CS. Quite many of the Finnish nouns used in CPs with Romani as the ML are modern terms lacking in Finnish Romani (2a), while others occurred as the speaker failed to think of the appropriate Romani noun (2b). Thirty of the noun switches are Finnish toponyms. Interestingly, Romani toponyms like *Baro fooros* 'Helsinki' are not found at all.

(2) a.   triin **kaksio** maan hin
         ('I have three, two-roomed flat')

    b.   jou ku šunjas douva, douva **tapahtuma**, jou rouvidas jakke buut
         ('when he heard about that that incident, he cried so much')

Note the hesitation *douva, douva* before the switch in (2b). Quite often, the speakers use phrases like *sar me phennaas* or *miten me phennaas* 'how would I say', when they fail to remember the approriate Romani item.
    The use of Finnish auxiliaries (such as *piti* 'had to' and *pittäisi* 'would have to') as in (3a) and inflected pronoun forms (such as *millä* 'with what', *mistään* 'from nowhere') instead of the corresponding Romani ones as in

(3b) may reflect a tendency to avoid complex structures and inflection para-digms in Romani. The Finnish Romani verb *mote* 'must' does not have a separate past tense form. Similarly, the Romani case system has been sub-ject to attrition (see Valtonen 1968: 169-71)

(3) a. ta douva **piti** aahhel tatto
   ('and it had to be hot')

  b. ku naas meen **millä** tšinjammas
   ('when we did not have what to buy with')

When Romani is the ML, mixed ML+EL constituents are more common (368 / 57.41 %) in bilingual CPs than EL islands (273 / 42.59 %). This is explained by the fact that most of the switches consist of undeclinable ad-verbs or nominative SG. forms of nouns with no case/number marking needed in either Romani or Finnish, inserted in positions determined by the ML grammatical frame (4a). Other cases indicate that the speakers tend to mix the grammars. Declined Finnish items tend to have the appropriate Fin-nish suffixes (4b), thus they tend to constitute EL islands. Examples like *me rikkavaa huolta* 'I take care' and *me džaa Helsinkiin* 'I go to Helsinki' show that the Finnish form chosen is one that follows the Finnish syntax. *Me rik-kavaa huolta* is corresponds to the Finnish utterance *minä pidän huolta*, and *me džaa Helsinkiin* to the Finnish utterance *minä menen Helsinkiin*.

Finnish nouns or verbs with Romani suffixes (as in 4c) or vice versa (4d) are very rare in the corpus. Valtonen (1968: 169-71) reports the bor-rowing of some Finnish cases (i.e. nominative PL., partitive PL., inessive SG., adessive SG. and allative SG.). Unlike he states, this seems to be mostly lim-ited to Finnish nouns inserted in Romani speech.

(4) a. khangari **ihan** neer doi hin
   ('the church is quite near there')

  b. liine Deevelesko **armoa**
   ('he received God's mercy')

  c. me **kant**otommas paani
   ('I carried water')

ko ame naa fattatommas toola **tykk**ite menna
('because we did not steal, they liked us')

d. na niekade maan ka rakkadommas kaalengo tšimb**ta**
('they did not forbid me to speak the Romani language')

In (4b), *armoa* is a Finnish partitive form of *armo* 'mercy'. The verb *kantotommas* in (4c) consists of the Finnish stem from *kant-aa* 'to carry' and the Romani verbal suffix *-otommas* (past tense, SG. 1st) . Similarly, *tykkite* combines the Finnish vernacular stem *tykk-* (from Finnish *tykätä* 'to like') and the Romani suffix for past tense SG. 3rd *-ite*. The noun *tšimbta* shown in (4d) is interesting as it is one of the few examples consisting of a Romani stem (*tšimb-* 'language') and a Finnish system morpheme. The suffix *-ta* is a Finnish partitive ending.

3. CS WITH FINNISH AS THE MATRIX LANGUAGE

Undoubtedly, all the Finnish Roma have today better language skills in Finnish than in Romani. In his old Ph.D. thesis, Vehmas (1961: 188) mentioned that for some 80 % of Roma families, the main or the only conversational language was Finnish. The functions of CS from Finnish to Romani are different than those we saw above. With the use of Romani items, the speakers probably wish to express solidarity with the group. Quite often, CS is also used to exclude *gadže* from a conversation.

CP length	Frequency	%
1	69	70.41
2	23	23.47
3	5	5.10
4	1	1.02

Table 3. Length and frequency of switches from Finnish to Romani.

Table (3) shows a pattern similar to the one presented in table (1). Here too, the majority of the switches are single-occurring lexemes.

Category	Frequency	%
Adjectives	3	4.35
Particles, adverbs	10	14.49
Nouns	29	42.03
Numerals	0	0.00
Pronouns	10	14.49
Verbs	17	24.64
Total	69	100.00

Table 4. Single occurring Romani lexemes classified according to their category.

Looking at the grammatical categories of the single-occurring lexemes (shown in table 4), we see a pattern different from the one found in the CS from to Romani to Finnish. Comparer to table (1), we notice, on the one hand, (i) the low number of Romani adverbs and on the other hand (ii) the high proportion of Romani content words (nouns ca. 42 %, verbs ca. 25 %) inserted. The independence of the two patterns presented in tables (1) and (5) could not be reliably tested using a Pearson Chi Square test, since 4 cells (33.3%) have an expected count of less than 5, and the minimum expected count is 0.50. It should also be noted that many of the instances where Romani nouns were found in CPs with Finnish as the ML have a metalinguistic character. One of the interviewees used Romani nouns when their meaning to the interviewer explaining in Finnish:

(5)   **ja** koi **on makkara**
         ('and *koi* is sausage')

The use of the Romani adjectives *tinalo* 'grazy' and *ilaka* 'nasty', found in otherwise Finnish CPs, exemplify well the function of CS to hide some of the meaning from outsiders:

(6)   **ykskin kääji ihan** tinalo
         ('one gadže woman [is] totally crazy')

se oli se vanha mies mahoton ilaka
('the old man he was terribly nasty')

When Finnish is the ML, mixed ML+EL constituents are much more rare (39 / 39.80 %) than EL islands (59 / 60.20 %). In most of the mixed ML+EL constituents, the Romani items are either undeclinable such as *alti* 'always', *ta* 'and', or nominative SG. forms of nouns, such as *Deevel* 'God', *hyösta* 'autumn', *däi* 'mother', *syy* 'needle', with no case/number marking needed. There are not many single lexical items with a Romani stem but a Finnish inflection. The most interesting examples are shown in (7):

(7)   **kaiken on** Deevel singla**nnut**
       ('Everything is blessed by God')

       entäs jos mustalaisnainen lokr**aa**
       ('what if a Gipsy woman fools')

The stem of the 2nd participle *singlannut* is from the Romani verb *singl-avaa* 'to bless', and the stem of the present tense SG. 3rd form *lokraa* from the Romani verb *lokr-avaa* 'to fool'.

The Romani EL islands vary considerably, from single lexemes, such as *siiva* 'gadže' (PL. NOM.), *Deeveles* 'God' (SG. ACC.) and *phennaas* 'I would say', to longer streches that constitute full NPs and VPs: *tšaȟȟesko moȟto* 'cabbage casserole', *džeelo butte skoola* 'went no more to school'.

4. BORROWING

While CS and borrowing are two different phenomena, they converge in many points. The single occurring CS lexemes and borrowed lexeme are often treated morphosyntactically in a very similar or even identical way in recipient language. Myers-Scotton (1997: 228) mentions three main differences between the two phenomena: (i) borrowed lexemes are accessible to monolingual speakers of the recipient language, while single occurring CS lexemes are not; (ii) borrowed lexemes constitute entries in the mental lexi-

con of both the recipient language and the source language; (iii) in lexical borrowing, lexemes from a sociopolitically more prestigious language are normally incorporated into a less commanding language.

Finnish Romani has borrowed many of its conjuctions from Finnish; the Finnish conjuctions are used by many speakers along with or instead of the original Romani ones. This type of lexical borrowing was already noticed by Thesleff (1901), who included the conjunction *että* 'that' in his dictionary. Valtonen (1968: 172) lists the following Finnish conjunctions borrowed:

	Finnish loan	Romani	Glos
adversative	mutta	bi	'but'
explanative	sillä	dooleske	'for'
causal	ko, ku	kaana	'because'
consecutive	että, nii että, jotta	ka	'so that'
final	että, jotta	te, at	'that'
conditional	jos	om	'if'
temporal	ko, ku	ka, kan, kaana	'when'
explicative	että, jotta	te, at	'so'
comparative	ku	sar	'as'

The corpus used for this study indicates that besides this, the Finnish coordinating conjunction *ja* 'and' is used along with *ta* 'and'. Table (5) illustrates the corpus frequencies of different conjunctions in the data with Romani as the ML. We see that the Finnish conjuctions have virtually replaced the original Romani ones at least in the speech of the interviewees.

Finnish Conj.	Freq.	%	Romani Conj.	Freq.	%	Total	Glos
ja	232	63.04	ta	136	39.96	368	'and'
että	78	97.50	at	2	2.50	80	'that'
jos	21	100.00	om	0	0.00	21	'if'
kun	137	88.96	ka	17	11.04	154	'when'
mutta	73	86.90	bi	11	13.10	84	'but'
Total	468	66.20		166	23.48	707	

Table 5. Finnish and Romani conjunctions in CPs with Romani as the ML.

In addition to the Finnish conjunction, Finnish Romani has borrowed the Finnish negation *ei* 'no'. The five types of negations found in the corpus are shown in (8):

(8)   nii manu naa naa ħunnela
        ('so Manu does not hear')

        peska tijaħħi, tši naa nas
        ('small shoes, nothing was')

        tši naa osuvaa **muuta** phennaa
        ('I cannot say anything else')

        ei naa dikjas peleški
        ('she did not read the cards')

        touva ei lustigo
        ('it [is] not funny')

The monolingual Romani types *naa*, *naa na(a)* and *naa nas* 'no' cover 77.5 % of the negations in the data where Romani is the ML. The mixed types *ei naa* and *ei nas* 'no' stand for 21.6 %. Bare Finnish *ei* is extremely rare (0.9 %).

Similarly, the nouns *kaalo* 'Roma' and *kaajo* 'non-Roma' are very frequently used by Roma in Finnish discourses. I this study, I have considered them borrowings from Romani.

5. CONCLUSIONS

To summarize: In this paper, I considered two kinds of intrasentential CS: (i) from Romani to Finnish and (ii) from Finnish to Romani. It was indicated that the two kinds of CS show fundamental differences as for both functions and patterns of CS.

CS from Romani to Finnish serves typically as a helping device for the speakers to fill linguistic gaps. The Romani lexicon has remained very small. The speech of the Finnish Roma has also shown attrition for a long time. As for the use of Romani items in discources in Finnish, it serves probably to express solidarity with the group. Quite often, CS is used to exclude *gadže* from a conversation.

There are far more switches from Romani to Finnish than vice versa. Most of the switches are single occurring lexemes. Streches longer than two words are rare. More than half of the single-lexeme-switches are different kinds of particles/adverbs, many conjunctions have been borrowed from Finnish. Other categories are generally less susceptible to CS: typically, Finnish content words are used if they are lacking in Finnish Romani (many modern terms), or if the speaker fails to remember them. There is also a tendency to use Finnish items to avoid complex Romani structures and inflection types. Mixed ML+EL constituents are more common than EL islands. Yet, declined forms of Finnish items nearly always to have the approriate Finnish suffixes.

Most instances of CS from Finnish to Romani, too, are single occuring lexemes. As for the grammatical categories of the single-occurring lexemes, we notice (i) the low number of Romani adverbs and (ii) the high proportion of Romani content inserted. Contrary to what we saw above, when Finnish is the ML, mixed ML+EL constituents are much less frequent than EL islands.

Lexical borrowings were discussed in short. In was shown that the Finnish conjuctions have virtually replaced the original Romani ones. Monolingual Romani types of negations are, however, more common than mixed or monolingual Finnish ones in the CPs with Romani as the ML.

REFERENCES

Cook, V. (1991) *Second Language Learning and Language Teaching*. Edward Arnold/ Hodder Headline Group: Melbourne.

Hedman, H. (1996) *Sar me sikjavaa romanes. Romanikielen kielioppiopas*. Jyväskylä: Opetushallitus.

Koivisto, V. (1987*) Rakkavaha romanes. Kaalengo tšimbako sikjibosko liin.* Helsinki: Ammattikasvatushallitus & Valtion painatuskeskus.

—— (1994) *Romano–finitiko–angliko laavesko liin. Roman–suomi–englanti sanakirja. Romany-Finnish-English Dictionary.* Kotimaisten kielten tutkimuskeskuksen julkaisuja 74. Helsinki: Valtion painatuskeskus.

Kopsa-Schön, T. (1996) *Kulttuuri-identiteetin jäljillä. Suomen romanien kulttuuri-identiteetistä 1980-luvun alussa. Kulttuuri-identiteetin jäljillä. Suomen romanien kulttuuri-identiteetistä 1980-luvun alussa.* Suomalaisen Kirjallisuuden Seuran julkaisuja 641. Tampere: Suomalaisen Kirjallisuuden Seura.

Myers-Scotton, C. (1993b*) Duelling Languages: Grammatical Structure in Codeswitching.* Oxford: Clarendon Press.

—— (1997) Code-switching. In Florian Coulmas (ed.), *The Handbook of Sociolinguistics*, 217–37. Oxford: Basil Blackwell.

Thesleff, A. (1901) *Wörterbuch des Dialekts der finnländischen Zigeuner.* Acta Societatis Scientiarum Fennicae XXIX:6. Helsinki.

Valtonen, P. (1968) *Suomen mustalaiskielen kehitys eri aikoina tehtyjen muistiinpanojen valossa.* Licentiate thesis. Helsinki.

Vehmas, R. (1961) *Suomen romaaniväestön ryhmäluonne ja akkulturoituminen.* Turun yliopiston julkaisuja sarja B 81. Turku: Turun yliopisto.