



**HAL**  
open science

## Unevenly mixed Romani languages

Evangelia Adamou, Kimmo Granqvist

► **To cite this version:**

Evangelia Adamou, Kimmo Granqvist. Unevenly mixed Romani languages. *International Journal of Bilingualism*, 2015, 19 (5), pp.525-547. 10.1177/1367006914524645 . halshs-01287045

**HAL Id: halshs-01287045**

**<https://shs.hal.science/halshs-01287045v1>**

Submitted on 21 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Unevenly mixed Romani languages

**Evangelia Adamou**

Oral Tradition Languages and Civilizations laboratory, French National Centre for Scientific Research (CNRS) & University Paris Sorbonne Nouvelle, France

**Kimmo Granqvist**

Department of Finnish, Finno-Ugrian and Scandinavian Studies, Helsingin Yliopisto, Finland

### **Abstract**

This study reports on language mixing in two Romani communities, with a century-long presence in Finland and in Greece respectively. A quantitative analysis of free-speech data shows that verbs from the contact languages, Finnish and Turkish, are systematically inserted into a dominant Romani speech with their respective Finnish and Turkish tense, mood, aspect, and person morphology. The insertion in language A of non-integrated single words from language B is atypical for classic code-switching and borrowing, but is a well-known mechanism in the creation of mixed languages. Unlike mixed languages, however, where no single dominant language can be identified, Romani is the main component in the corpora under study. We suggest that this type of Romani language mixing illustrates an early stage of mixed language formation that did not develop into an independent mixed language, owing to changes in the sociopolitical settings.

**Key-words:** Code-switching; mixed languages; Romani; bilingual speech

### **1. INTRODUCTION**

The present study is an analysis of language mixing in two Romani communities, one of which settled in Finland and the other in Greece, with a century-long presence in each country. The Finnish and the Greek Thrace Romani communities reported on here have been engaged in the process of shifting to their respective contact languages, Finnish and Turkish, since the nineteenth century. A quantitative look at two free-speech corpora (totaling 18,800 word tokens) shows that the current contact language verbs, representing 10-12% of the verbs in the corpus, always retain the verb morphology of the contact language when inserted into a dominant Romani environment, as shown in (1) and (2) for person and tense, mood, aspect (TMA) markers. In addition, in the

Romani-Finnish corpus, Finnish nouns keep their Finnish nominal morphology, namely, case, shown in (2), and number.

Thrace Romani (Greece) < Romani (in plain), Turkish (in bold), Greek (underlined)

- (1) latʃo afu gadal **dyfym-ijor-sənəs** te dzavtar mange  
good since this\_way think-PROG-2PL COMP go.1SG.DIR me.DAT  
“Fine, since this is what you think, I’ll leave.”  
(Speaker 9, excerpt from the tale “The coward and the giants”, sentence 95,  
available online at <http://lacito.vjf.cnrs.fr/pangloss/index.htm>)

Finnish Romani (Finland) < Romani (in plain), Finnish (in bold)

- (2) a. **ja** do:ri **rukoil-i-n** de:vel-es-ta  
and there pray-PST-1SG god-OBL.SG-ABL  
“And there I prayed to God.”
- b. maŋ-jom de:vel-es ta de:vel l-i:jas ma:n **oma-ks**  
pray-PST.1SG god-OBL.SG and god take-PST.3SG me own-TRN  
“I prayed to God and God took me to his own.”
- c. li:jas **kokona:n** ma:n **oma-ksi** **ja** de:vel **täytt-i**  
take.PST.3SG entirely me own-TRN and god fill-PST.3SG  
“God took me wholly unto himself and filled me.” (Speaker 1)

The insertion of morphologically-intact elements from the contact languages is rare in code-switching and incompatible with most definitions of borrowing (Myers-Scotton 2002, Poplack and Dion 2012). It is nevertheless a process typical of mixed languages, such as Michif (Bakker 1997a), Light Warlpiri (O’Shannessy 2013), and Gurindji Kriol (Meakins 2012). But unlike evenly mixed languages, for which a single dominant language cannot be identified, the Romani corpora under study have a clear dominant Romani component. Rather, the Romani data resemble code-switching in the early stages of the formation of mixed languages (O’Shannessy 2012, 2013, McConvell and Meakins 2005, Meakins 2012) giving rise to what we call here “unevenly mixed languages.”

Similarly to what has been described for mixed languages, the Romani mixing developed in highly bilingual communities, due to conflicting processes of language shift and language maintenance. Following Myers-Scotton (2013), we suggest that this type of unevenly mixed Romani language results from an arrested Matrix Language Turnover (Myers-Scotton 2002), which probably took place during the nineteenth century. This particular language mixing was then transmitted from one generation to another, providing an example of a challenging, new type of language mixing.

## 2. THEORETICAL BACKGROUND

“Mixed languages are the result of the fusion of two identifiable source languages, normally in situations of community bilingualism” (Meakins 2013: 159). There are at least two types of documented mixed languages: those which draw the grammar from one language and the lexicon from another, like Ma’á and Angloromani (Bakker and Mous 1994, Matras 2010), also known as G(rammar)-L(exicon) mixed languages (Meakins 2013); and mixed languages that show compartmentalization of elements of the two languages, or V(erb)-N(oun) mixed languages (Meakins 2013), like Michif (Bakker 1997a), Light Warlpiri (O’Shannessy 2013) and Gurindji Kriol (Meakins 2012).

The process of mixed language formation has long been debated, but today it is acknowledged that mixed languages may result from general processes observed in most language contact situations such as code-switching (Myers-Scotton 2002, Bakker and Matras 2013, Meakins 2013). Based on the analysis of the productions of older and younger speakers, McConvell and Meakins (2005) and O’Shannessy (2012) provide evidence supporting the creation of mixed languages following several stages of code-switching. O’Shannessy (2012: 330-31) argues that code-switching, characterized by the possibility of switching at various points, at some stage becomes a fused lect (FL), a form of stabilized code-switching, defined by Auer (1998) as follows:

[W]hile LM (Language Mixing) by definition allows variation (languages may be juxtaposed, but they need not be), the use of one “language” or the other for certain constituents is obligatory in FLs; it is part of their grammar, and speakers have no choice. (Auer 1998: 15).

A step further in the development of an independent mixed language is the creation of innovative patterns that do not exist in the source languages (O’Shannessy 2012, 2013, Meakins 2012, 2013).

A similar view, which takes code-switching as the basis of mixed languages, is expressed by Myers-Scotton (2002) within the Matrix Language Frame model (MLF) (Myers-Scotton 1993). In the MLF model, the language that sets out the grammatical frame is the Matrix Language (ML), while the other participating language is called the Embedded Language (EL). The grammatical frame is defined by the morpheme order and the system morphemes (inflections, system words), which contrast with the content morphemes (nouns, adjectives, verbs). Myers-Scotton suggests that mixed languages result from an arrested Matrix Language Turnover, following a stage of “composite codeswitching”, during which there is more than one source for the Matrix Language (Myers-Scotton 2002: 105, 269).

### 3. BACKGROUND ON ROMANI

3.1. HISTORICAL AND DIALECTOLOGICAL OVERVIEW. Romani is an Indo-Aryan language spoken throughout Europe, in the Americas, and in Australia. The earliest migrants, who probably belonged to service-providing castes, arrived from India during the period of the Byzantine Empire, around the tenth century. During this period, Romani was considerably influenced by Greek. At the end of the Byzantine era, several groups migrated toward western and northern Europe, and new contact languages were added.

In contemporary Romani linguistics, dialectal categorization based on the geographical spread of linguistic features has become dominant (Matras 2005, 2010). Matras (2005, 2010) identifies a cluster of isoglosses dividing the dialects of western and northern Europe from those of southeastern Europe. Four main dialectal groups are distinguished: Northern (subdivided into Northwest and Northeast), Central, Vlax, and Balkan Romani. The Thrace Romani variety spoken in Greece belongs to the Southern Vlax group, whereas the Finnish Romani variety belongs to the Northwestern group; the speakers recorded speak the most widespread Eastern sub-dialect of Finnish Romani (Valtonen 1968: 246-250; Koivisto 1987; Hedman 1996).

3.2. ROMANI IN CONTACT AND LANGUAGE ENDANGERMENT. In a great variety of contact settings, Romani speakers have been multilingual. Romani varieties have generally integrated lexical and structural borrowings from the contact languages (Matras 2002, Elšík and Matras 2006). Language contact in Romani communities has always been asymmetric, owing to the low prestige of the Romani language compared to the languages spoken by the majority populations with whom the Romani communities are in contact (Friedman 2000). In some settings, a language shift to the dominant majority languages has been generalized, i.e., in the traveler communities of Britain, Wales, and the Scandinavian countries, as well as in the Spanish and Portuguese Romani communities. Nevertheless, in some cases, vocabulary and certain fossilized structures of Romani origin have been mixed with the morpho-syntactic frame of the majority languages, giving rise to what is often known as Para-Romani varieties, used as a means of in-group communication, e.g., Angloromani, Scandoromani, and Caló (Cortiade 1991; Bakker 1995; Matras 2010; Carling, Lindell, and Ambrazaitis, in press).

Finnish Romani belongs to category C on the language endangerment scale, being “severely endangered,” with only the grandparental generation having fluency (Krauss 2006). According to Kopsa-Schön (1996: 44), in the 1990s only the elderly Roma (those over about 65 years of age) were able to communicate fluently in Romani. The shift is taking place toward Finnish, and younger Romani speakers may still acquire a mixed variety of Romani, although not as their first language, but as their second language,

acquired in adolescence during their socialization process within the Romani community of Finland (Thesleff 1899: 472; Vuorela and Borin 1998: 60; Borin 2000: 75).

Thrace Romani can be qualified as “unstable, eroded,” rated A– in Krauss’s scale of language endangerment (Krauss 2006). This category holds for situations in which the language is spoken in some localities by children, but a shift at the level of the parental generation is taking place in other localities or in families within one locality. The language shift for the Muslim Roma of Greek Thrace is taking place toward Turkish, which is the language of the Muslim minority in the area, rather than to Greek, which is the language of the state and administration (Adamou 2010).

#### 4. DATA

The type of Romani phenomena on which we are reporting here have been discussed in the contact literature in a corpus-illustrated approach: for instance, Rusakov (2001) for North Russian Romani; Elšík and Matras (2006) for Crimean Romani; Granqvist (2007) for Finnish Romani; and Adamou (2010) and Friedman (2010, 2013) for Romani spoken in the post-Ottoman area of the Balkans. The present study is the first attempt to account for Romani language mixing with a quantitative look at free-speech data. The data that are systematically analyzed in this paper are part of larger corpora that are either untranscribed or transcribed, but not tagged for contact phenomena.

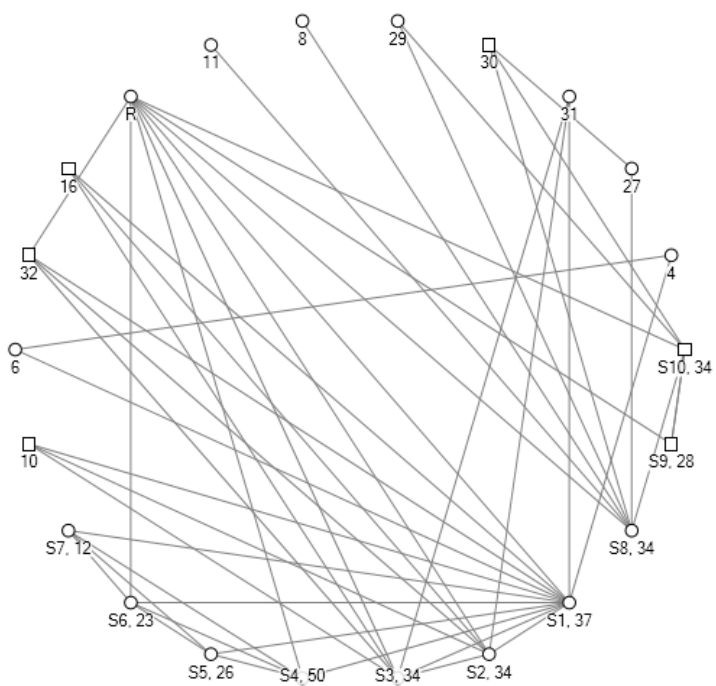
4.1. THRACE ROMANI CORPUS. The Thrace Romani corpus includes data from storytelling, interviews with the first author of the paper, and in-group conversations among 21 Roma speakers in the presence of the researcher. The data were collected during four fieldwork visits carried out between 2007 and 2010. The recordings took place in the house, yard, or workplace of the participants. Along with the main group of speakers, several friends and family members regularly stopped by, as is frequently the case in Romani communities. Although the verbal interactions of all participants were transcribed, the participation of the speakers varied considerably (see Figure 1). All speakers had little or no formal schooling in Turkish or Greek and were not literate in Romani. Most of them grew up in Thrace, but had also spent several years in other Greek cities for work purposes. The resulting corpus for Thrace Romani has 5,816 word tokens.

The Thrace Romani data have been transcribed by the first author of the paper with a specifically developed tool for multilingual corpora *Jaxe*<sup>1</sup> which allows tagging the

---

<sup>1</sup> Developed by Pascal Vaillant for the program ‘Towards a multi-level, typological and computer-assisted analysis of contact-induced language change’, PI: Isabelle Léglise (CNRS), funded by the French Research Agency (ANR), 2010-2014.

constituents with regard to the languages in contact and other sociolinguistic information. Each word was tagged as Romani, including all Indic words and borrowings from previous language contact settings or the current contact language word, namely, Turkish or Greek. The category “multiple” was added for cases in which the word could be both Turkish and Greek. All words, from Romani or the current contact language, were also tagged, depending on the word class, as nouns, verbs, adjectives, adverbs, conjunctions (e.g. complementizers, subordinators), particles, and determiners. The resulting corpus can be interrogated through a concordance tool, which is currently being developed.<sup>2</sup>

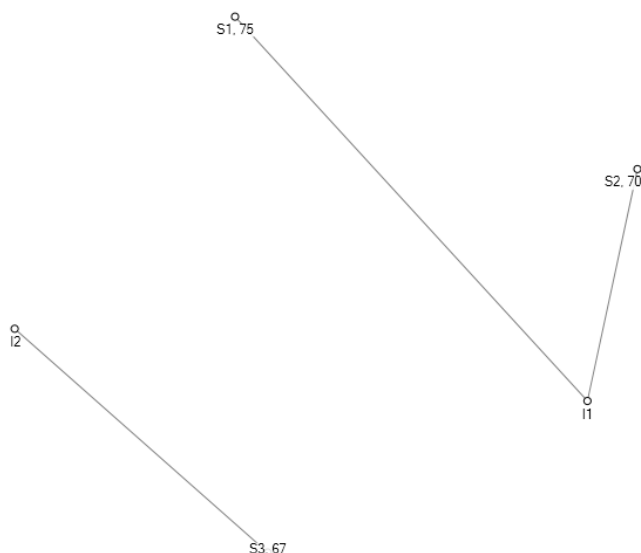


**Figure 1.** Verbal interactions in the Thrace Romani corpus (participants’ age and sex are coded; circles represent female participants and squares, male participants; R = the researcher; S = the speakers who participated the most in the discussions).

4.2. FINNISH ROMANI CORPUS. The Finnish Romani corpus is based on three separate conversations between in-group interviewers and three fluent Romani speakers. The recordings took place in 1995 during a Romani language seminar in what can be described as a semi-formal setting. The conversations were carried out during three one-

<sup>2</sup> Developed by Anne Fernandez-Garcia for the program ‘Multilingual corpora’, PI: Isabelle Léglise (CNRS), of the French Excellence Laboratory ‘Empirical Foundations of Languages’.

hour sessions between an interviewer and a speaker. The interviewers were two Roma women in their 30s, both proficient speakers of Finnish Romani. The three elderly Roma interviewees were born in 1920, 1925, and 1928. While all three were proficient in Finnish Romani, Finnish was the dominant language in their everyday interactions. The resulting corpus, including the data from the interviewers and the older speakers, consists of 20,111 word tokens. Nevertheless, the interviewers' data are not analyzed in detail because they result from consistently planned, Finnish-Romani speech, intended to elicit Finnish-Romani responses. The data analyzed from the three older speakers constitute a corpus of 13,019 word tokens, which serves as the basis for our analysis. The Finnish corpus was transcribed in Word format by a trained community member, Hellevi Hedman-Valentin. The clauses were tagged by the second author of the paper with respect to the dominant language and the switching points to Finnish or Romani. In addition, each word was tagged for word class. The analysis was then conducted using Excel.



**Figure 2.** Verbal interactions in the Finnish Romani/Finnish corpus (participants' age and sex are coded; circles indicate female participants; I = the in-group interviewers; S = the speakers).

## 5. RESULTS

In order to understand the type of mixing that occurs in the Romani corpora under study, we attempted to address, from a quantitative perspective, several criteria which have been discussed in the literature on mixed language creation (for an overview see Meakins 2013). In 5.1 we analyze, quantitatively, the corpora with respect to the source languages. In 5.2 we examine the length of switches. In 5.3 we examine the functional



role of the morphemes from each source language based on an analysis of word classes. In 5.4 we take a close look at verb and noun morphology. Lastly, in 5.5 we examine inter-speaker variation with a special focus on verbs and nouns. 5.6. briefly summarizes these results.

5.1. THE NUMERICALLY DOMINANT LANGUAGE. One of the criteria suggested in order to identify the matrix language (ML) of a bilingual corpus is the criterion of frequency, according to which “the ML is the language of more morphemes in interaction types including intrasentential code-switching” (Myers-Scotton 1993: 68). This criterion was later rejected by Myers-Scotton (2002: 61) who favors the System Morpheme Principle. Moreover, as mentioned by Muysken (2000: 66), the morpheme-frequency criterion should be taken into consideration in combination with the typology of the languages in contact.

We agree that the quantitative approach alone is not sufficient and should be combined with other parameters. Nevertheless, a look at the overall composition of a multilingual corpus gives a general idea of the proportion of word tokens that are shared with the current contact languages, i.e. Finnish or Turkish and Greek, as compared to those not shared, i.e., the Indo-Aryan lexicon and the numerous integrated borrowings from earlier contact languages, including Persian, Armenian, and Byzantine-era Greek. Specifically for Finnish Romani, Swedish<sup>3</sup> and other Germanic languages exerted a considerable influence in the past, as have Romanian and the Slavic languages for Thrace Romani. To determine the proportion of word tokens, we counted all the words shared with the current contact languages and calculated the percentage they represent in the total number of words in the corpus.

The analysis shows that 80% of the Thrace Romani corpus is composed of Romani word tokens (including word tokens from the past contact languages); 15% of word tokens are shared with Turkish, and 4% with Modern Greek, while 1% of the words could be either Turkish or Greek, designated as “multiple” (see Table 1). This is shown in (3) for Thrace Romani.

Thrace Romani (Greece) < Romani (in plain), Turkish (in bold), Greek (underlined)

(3) a.   kon                   **alna-dzak-lar**  
          who                 understand-FUT-3PL  
          “Whoever will understand,”

          b.   ka         dav         leske         me         tʃa  
              FUT       give.1SG   3SG.DAT   my.OBL   daughter.OBL  
              “I’ll give him my daughter.”

---

<sup>3</sup> From the twelfth century until the early nineteenth century, Finland was part of Sweden.

(Speaker 10, excerpt from the tale “The louse and the Rom”, sentence 15, available online at <http://lacito.vjf.cnrs.fr/pangloss/index.htm>)

**Table 1.** Thrace Romani corpus: Distribution of word tokens per language

	<b>Romani</b>	<b>Turkish</b>	<b>Greek</b>	<b>Multiple</b>	<b>Total</b>
<b>%</b>	80%	15%	4%	1%	100%
<b>Tokens</b>	4661	876	253	26	5816

In the Thrace Romani corpus, clauses with Turkish as the numerically dominant language are rare and are generally participant-related, i.e., triggered by the interactions with monolingual speakers of Turkish and Greek, with Romani speakers who have shifted to Turkish or Greek, as well as in reported speech. Below is an example of a female speaker who addresses her friend in Turkish (both women live in another neighborhood and have shifted to using Turkish in daily life); she immediately formulates the question in Romani when addressing a young girl who lives in the neighborhood where Romani is still (at least partly) transmitted:

Turkish (in bold), Romani (in plain), Multiple (underlined)

Speaker 1 to her friend:

- (4) a. **yzgjanən**      **kəzə**      **dilmi**      **bu**      mar  
 NP.GEN              girl.POSS      NEG.QUEST      this      INTJ  
 “Hey, isn’t she Yzgjan’s daughter?”

Speaker 1 to the girl:

- b. yzgjanaki      i      tʃei      naj      san      tʃe  
 NP.GEN              DEF.F      daughter      be.NEG      be.2SG      INTJ  
 “Hey, aren’t you Yzgjan’s daughter?”

The Finnish corpus shows a majority of Romani words, 65%, with 35% Finnish words (see Table 2).

**Table 2.** Finnish Romani corpus: Distribution of word tokens per language

	<b>Romani</b>	<b>Finnish</b>	<b>Total</b>
<b>%</b>	65%	35%	100%
<b>Tokens</b>	8410	4621	13031

The Finnish Romani corpus can be further sub-divided into clauses with a majority of Romani tokens and clauses with a majority of Finnish tokens. Clauses with Finnish as the numerically-dominant language amount to 20% of the total or 1,037 clauses. Finnish-dominant clauses, such as in the example (5), are not participant related. In

these clauses, Finnish provides the majority of words (96%) with only 4% of Finnish Romani tokens (Table 3).

Finnish Romani (Finland) < Romani (in plain), Finnish (in bold)

- (5) **ihan** bar-o tše:r **sem:onen** **taval:inen** **ni:ŋku** **ma:laistalo**  
 quite big-M.SG house such usual like farm\_house  
 “Quite a big house, a common one, like a farm house.” (Speaker 1)

**Table 3.** Finnish Romani corpus: Distribution of word tokens per language in the Finnish dominant clauses

	<b>Romani</b>	<b>Finnish</b>	<b>Total</b>
<b>%</b>	4%	96%	100%
<b>Tokens</b>	119	2689	2808

Clauses with Finnish Romani as the numerically dominant language are the most frequent in the corpus, representing 80% of the entire corpus. In these clauses, Finnish Romani provides the majority of tokens, i.e. 81% of Romani words as opposed to 19% of Finnish words (Table 4), as shown in (6) and in (2).

Finnish Romani (Finland) < Romani (in plain), Finnish (in bold)

- (6) a. hin tsonak br:al tsonatik-a aŋrusti-a  
 be.3SG gold on golden-PL ring-PL  
 “Wears golden rings,”
- b. prinp-os-ko sy: **kame:** **neula** anrus **ja** ka:neŋ:ier-e  
 breast-OBL.SG-GEN needle cameo needle ring and earing-PL  
 “a brooch, cameo brooch, and earrings.” (Speaker 3)

**Table 4.** Finnish Romani corpus: Distribution of word tokens per language in the clauses with Finnish Romani as the dominant language

	<b>Romani</b>	<b>Finnish</b>	<b>Total</b>
<b>%</b>	81%	19%	100%
<b>Tokens</b>	8410	1934	10344

To conclude, the comparison of the two Romani corpora indicates that, in both cases, Romani is the dominant language of the corpus. Nevertheless, the Finnish Romani corpus shows 15% fewer Romani words than the Thrace Romani corpus. This percentage correlates with the more advanced stage of language shift in the Finnish Romani community, where, despite the semi-formal setting and the interviewer’s incentives to use Romani, Finnish is abundantly present in the production. By contrast,

even though the Thrace Romani corpus is more spontaneous, it still shows more Romani words than the Finnish corpus.

If we compare the Thrace Romani corpus with the sub-corpus of the clauses with numerically dominant Finnish Romani, we observe that the proportion of Romani tokens is very similar in the two corpora: 81% of Romani words for Finnish Romani and 80% of Romani words in Thrace Romani. Nevertheless, the Finnish words alone represent the proportion taken up by both Turkish and Greek tokens in the Thrace Romani corpus. Therefore, the impact of Finnish should be understood as being much greater on Finnish Romani than is the impact of Turkish on Thrace Romani.

5.2. LENGTH OF SWITCHES. Code-switching can be “insertional” (Muysken 2000), a type also known in other studies as “intrasentential” (Myers-Scotton 1993). Myers-Scotton (1993) defines intrasentential code-switching as containing at least one word from the embedded language (EL) and any number of words from the matrix language (ML). When code-switching occurs at the boundary of a clause, it is called “alternational”, covering what is sometimes also known as extrasentential and intersentential code-switching. In order to evaluate the main type of switching, we examine first one of the diagnostic features suggested in Muysken (2000: 230), namely the length of the insertions from language A in a numerically-dominant clause from language B (nested A B A). Word class (content word vs. adverb, conjunction) and switch site (e.g. peripherality) are examined separately in the section 5.3. dedicated to the study of word class.

Single-word switching (of any word class) is the most common type found in the Thrace Romani corpus, with 74% of the Turkish words and 88% of the Greek words being single-word tokens (see Table 5).

**Table 5.** Thrace Romani: Length of Turkish and Greek word-tokens

	Turkish		Greek	
	tokens	%	tokens	%
<b>1 word-token</b>	379	74	60	88
<b>2 word-tokens</b>	83	16	2	3
<b>3 word-tokens</b>	42	8	2	3
<b>&gt; 4 word-tokens</b>	12	2	4	6
<b>Total</b>	516	100	68	100

In the Finnish Romani corpus, clauses with Romani as the dominant language also show a majority of Finnish single-word insertions (75%) followed by 16% of two-word tokens and 6% of switches of three-word tokens (see Table 6). This is also the case in clauses with Finnish as the dominant language with 65% of single-word insertions (see Table 7).

**Table 6.** Finnish Romani corpus: Length of Finnish word tokens in mixed clauses with Romani as the dominant language

	<b>Finnish</b>	
	<b>tokens</b>	<b>%</b>
<b>1 word-token</b>	1036	75%
<b>2 word-tokens</b>	216	16%
<b>3 word-tokens</b>	80	6%
<b>&gt; 4 word-tokens</b>	46	3%
<b>Total</b>	1378	100%

**Table 7.** Finnish Romani corpus: Length of Romani word tokens in mixed clauses with Finnish as the dominant language

	<b>Romani</b>	
	<b>tokens</b>	<b>%</b>
<b>1 word-token</b>	26	65%
<b>2 word-tokens</b>	11	27,5%
<b>3 word-tokens</b>	3	7,5%
<b>Total</b>	40	100%

To summarize, both corpora show a majority of short switches from language A in a dominant clause from language B, as is typical in insertional switching (Muysken 2000).

5.3. WORD CLASSES. A word-count allows for an initial evaluation of the corpus but needs to be combined with other criteria, such as the functional distribution of the words in the corpus. Indeed, a bilingual corpus with a great number of contact language nouns is clearly distinct from a corpus with a great number of contact language grammatical morphemes. This is expressed in the MLF model of code-switching by the System Morpheme Principle (Myers-Scotton 1997: 98), according to which the matrix language is identified based on the contribution of the system morphemes. An analysis of the numerically Romani-dominant clauses is presented here in order to identify the proportion of the various word classes.

Figure 1 presents the distributional analysis of word tokens in the Thrace Romani corpus per language based on word classes. The analysis of the Thrace Romani corpus shows 25% of Turkish nouns and 10% of Modern Greek nouns (only 2% of multiple) or a total of 37%. When it comes to verbs, the Thrace Romani variety shows a relatively high number of tokens from the current contact languages, namely, 12% of Turkish verbs and 2% of Greek verbs. Adverbs are also frequently Turkish (29%) and Greek (6%). We also note 19% Turkish and 9% Greek adjectives. Last, the Thrace Romani corpus shows a mixture of conjunctions with 66% of Romani conjunctions (mainly *kaj* used as a relative marker, and *te* as a complementizer), 23% Turkish, and 11% Greek. The word class that is least affected by contact is pronouns (personal, interrogative, and

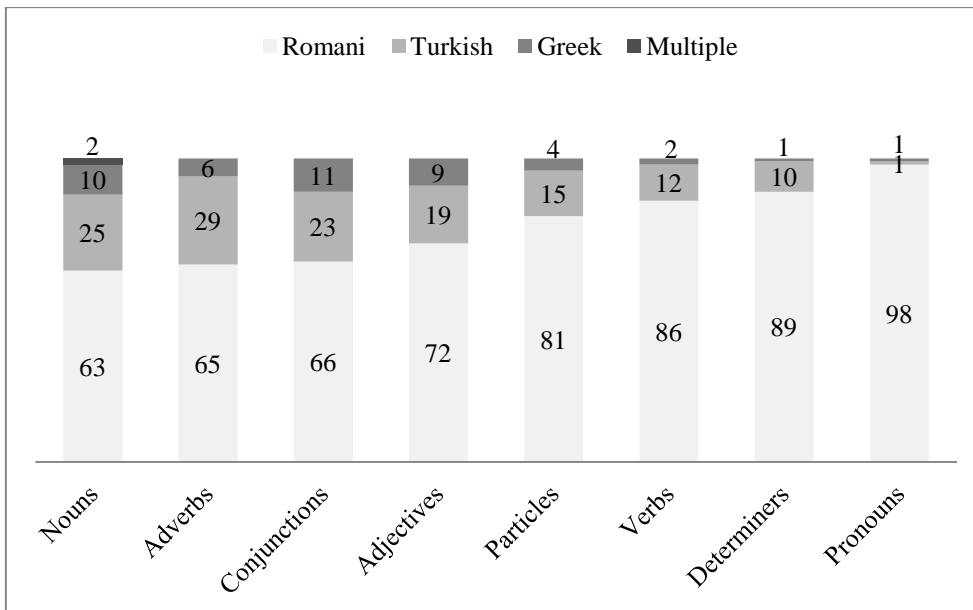
indefinite pronouns), i.e., 98% of Romani pronouns. Typologically, the fact that Turkish, like Romani and Greek, is a pro-drop language may have an effect on this result. In sum, what characterizes a Thrace Romani utterance is the lack of some Turkish word classes such as pronouns and determiners.

The clauses with dominant Finnish Romani show 72% of Finnish conjunctions, 27% of Finnish particles, 22% of Finnish adverbs, and 13% of Finnish nouns (see Figure 2). As opposed to the Thrace Romani corpus, which shows practically no Turkish pronouns, Finnish Romani also integrates 5% of Finnish pronouns (personal, interrogative, demonstrative, reflexive, possessive, and indefinite pronouns). Moreover, the proportion of verbs from Finnish<sup>4</sup> is similar to that of Turkish verbs in the Thrace Romani corpus, i.e., 10%.

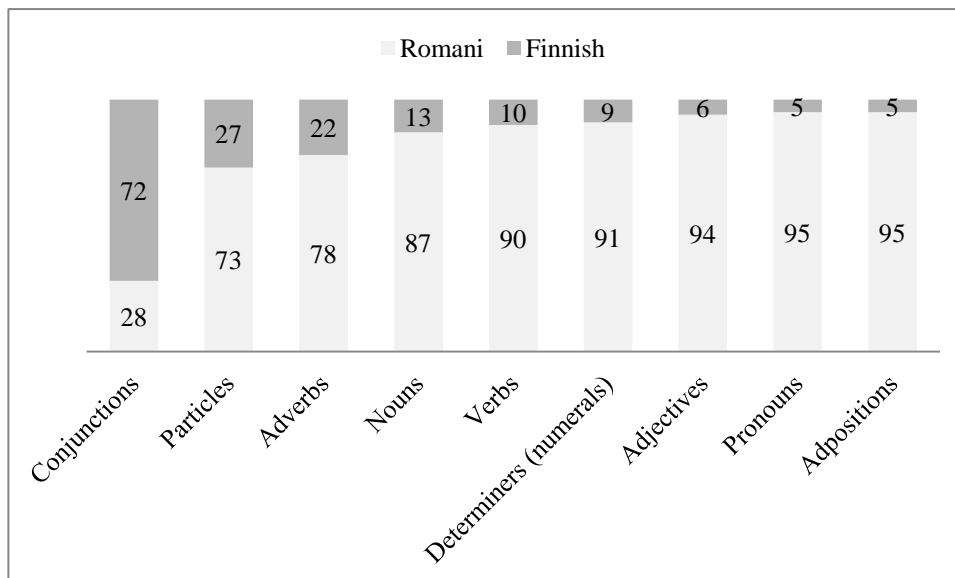
To summarize, in the two corpora conjunctions and adverbs are among the most frequent current-contact language items and could be analyzed in Muysken's terminology as alternational switches. The frequency of contact-language conjunctions is also well known in the literature on borrowing (Matras 2009). The two corpora also show that nouns are very frequent, a result which is also in accordance with studies on code-switching and borrowing. It is interesting to note that in the Thrace Romani corpus, where short switches are favored, nouns, adverbs and conjunctions represent from 34-37% of their word class. In Finnish Romani in contrast, where speakers frequently alternate between Romani-dominant and Finnish-dominant clauses, the dominant Romani speech is characterized by a majority of Finnish conjunctions whereas all the other word classes are less affected by Finnish. Lastly, in both corpora pronouns are among the least frequent word classes, a result which may also be related to the fact that all languages in contact are pro-drop languages.

---

<sup>4</sup> The Finnish negation verb *ei* has been systematically regarded as a verb, since it retains its Finnish verbal inflection, even though it sometimes behaves like a negation participle analogous to the Romani negation particle *naa*.



**Figure 1.** Distribution of word tokens per language and word class in Thrace Romani (%).



**Figure 2.** Distribution of word tokens per language and word class in dominant Finnish Romani speech (%).

5.4. MORPHOLOGY. Several authors consider that a definitional criterion of the matrix (or base) language is the main verb of a clause and its inflection (Treffers-Daller 1994, Muysken 2000, Meakins 2013). According to Myers-Scotton (2002: 61), verb inflection

should be a parameter examined in relation to all other system morphemes, as expressed by the System Morpheme Principle (Myers-Scotton 1997: 98). We examine in 5.4.1 and in 5.4.2 how verbs and nouns from the contact languages are (not) integrated in terms of morphology in a dominant Romani speech.

5.4.1. VERB MORPHOLOGY. In Thrace Romani, Turkish verbs consistently take Turkish verb morphology, i.e., person markers, negation, causative, and TMA markers, meaning preterit, progressive, future, and partly optative; the Turkish evidential morpheme was reanalyzed as an adverb with the Greek meaning of reporting the truth value of a statement (Adamou 2010, 2012). In the Thrace Romani corpus, we can attest to a variety of Turkish verbs, i.e., motion (*koyul-* “to approach”) and posture verbs (*uzan-* “to lie”), perception-cognition verbs (*düşün-* “to think,” *alna-* “to understand,” *konuş-* “to talk”), emotion verbs (*begen-* “to like,” *aci-* “to pity”), and several action verbs (*oku-* “to read,” *yaz-* “to write”). Turkish phonology is generally respected, including non-inherited phonemes accompanying the borrowed item, such as /y/, /ø/, and /w/. Phonological adaptation might take place in some cases, i.e., the metathesis of /nl/ to /ln/, as in the Turkish verb *anlayacaklar* > Thrace Romani [alnadzaklar], meaning “they will understand”, illustrated in example (3a). Contrary to the Turkish verb-final canonical order, the borrowed verbs follow the Romani verb-initial word order, although topicalization and focalization result in frequent SV-OV structures (Arvaniti and Adamou 2011).

The Romani equivalent of the Turkish verbs is generally not used when speaking Thrace Romani; for example the verb *konuş-* “to talk” is used in our corpus consistently by all speakers (7a). Nevertheless, knowledge of the Romani equivalents is obtained through frequent contact with speakers of other Romani varieties with just a limited number of Turkish verbs borrowed during the Ottoman times (Adamou 2010). The use of the Romani form is thus not excluded, as shown in (7b) and (7c) for the verb “to talk” used by the same speaker. The limited corpus we are working on doesn’t allow for a systematic study of the Romani equivalents of single Turkish verbs and a different, probabilistic approach is suggested in 5.5.

Thrace Romani < Romani (in plain), Turkish (in bold), Multiple (underlined)

- (7) a. dʒan-es kasa **konuʃ-ijor-sun** akana  
 know-2SG who.INSTR talk-PROG-2SG now  
 “Do you know who you’re talking to now?” (Speaker 8)
- b. amen muruʃ tʃavo naj hajde **konuʃ**  
 1PL.OBL male boy be.NEG.3SG INTJ talk.IMP.2SG  
 “We don’t have a boy (in the family)...come on, talk!” (Speaker 4)



- c. **airsəs ep** me ka orbisarav  
 useless all\_the\_time 1SG.NOM FUT talk.1SG  
 “Useless (girl)! Am I to do the talking all the time?” (Speaker 4)

It is important to note that Romani words tend to precede the Turkish verbs (see Table 8) and that no flagging is observed, i.e., pauses, hesitation phenomena, repetition, metalinguistic commentary (Muysken 2000). Moreover, a clear majority of Romani words follows the Turkish verbs, i.e. 80% (Table 9).

**Table 8.** Words preceding Turkish verbs

	Romani word	Turkish word	Total	
<b>Turkish verb</b>	68%	32%	100%	<b>%</b>
	102	48	150	<b>tokens</b>

**Table 9.** Words following Turkish verbs

	Romani word	Turkish word	Total	
<b>Turkish verb</b>	80%	20%	100%	<b>%</b>
	123	30	153	<b>tokens</b>

Finnish verbs are integrated with Finnish verb morphology, including person, mood, and tense marking. They may have a Romani equivalent, as shown in (2), for the verb “pray”, or not. Finnish participles, shown in (8b), and infinitives are equally used, e.g., *puhumas:a* - “speaking”, as well as Finnish negation, either in variation with the Romani negation – *na:*, *na: na(a)* and *na: nas* – or co-occurring with it.

Finnish Romani < Romani (in plain), Finnish (in bold)

- (8) a. sa:r-o hin tʃiʃk-as  
 everything-M.SG be.3SG good-ADV  
 “Everything is well”

- b. **ja** sa:r-o hin de:vel **mu:t:a-nut**  
 and everything-M.SG be.3SG god change-ACT.PST.PTCP  
 “and God has changed everything.” (Speaker 1)

Finnish verbs with Romani morphology are very rare in the corpus (14 tokens out of 225). They mainly occur for the verb *tykätä*, “to like,” adapted into Romani as *tykkuv-*, and more rarely for the verb *kantav-*, “to carry,” shown in (9).

- (9) me **kant-**otom:as pa:ni  
 1SG carry-PST.1SG water.NOM  
 “I carried water.” (Speaker 2)

The phonology of colloquial Finnish is generally retained, including general gemination (e.g. *pitäisi* > *pit:äisi* “should”) and vowel harmony, which Finnish Romani has partly adopted into its phonology. The Finnish imperatives, *kato* – “look!” and *ku:le* – “hear!”, are frequent and are used in Romani speech, as in Finnish, as tags (Kovanen 2010). Likewise, the Finnish copula/auxiliary *ol:a* “to be” is frequently used to form Romani analytical past tenses.

In the Finnish Romani-dominant clauses, Finnish verbs are preceded by a Romani word in 78% of cases and followed by a Romani word in 65% of cases (Tables 10 and 11).

**Table 10.** Words preceding Finnish verbs

	<b>Romani word</b>	<b>Finnish word</b>	<b>Total</b>	
<b>Finnish verb</b>	78%	22%	100%	<b>%</b>
	147	41	188	<b>tokens</b>

**Table 11.** Words following Finnish verbs

	<b>Romani word</b>	<b>Finnish word</b>	<b>Total</b>	
<b>Finnish verb</b>	65%	35%	100%	<b>%</b>
	146	79	225	<b>tokens</b>

To conclude, tense, mood, aspect, and person markers accompany the Turkish and Finnish verbs. If we chose to follow the criterion of identification of the matrix language based on the verb and its morphology, then these clauses would have to be considered as Turkish and Finnish clauses respectively, illustrating a matrix language turnover. Nevertheless, a quantitative look at the corpora shows that contact-language verb morphology occurs in a numerically-dominant Romani speech, at the level of the entire corpus and of the clause, when applicable.

5.4.1. NOMINAL MORPHOLOGY. In the Thrace Romani corpus, the Turkish nominals take up Romani morphology, such as case and gender marking (distributed according to the word’s ending in the donor language). In contrast, in the Finnish Romani corpus, the Finnish nouns are always inserted into a dominant Romani speech with the Finnish case marking, as shown in (2b) for the translative case and in (10) for the partitive case. Examples in all Finnish cases can be found, i.e. accusative as in *rakkiboskero-n* ‘speaker-ACC.SG’, inessive as in *salmi-i-s* ‘Salmi-INESS’, adessive as in *kaale-i-lla* ‘Roma-PL-ADESS’, allative as in *gravjöre-i-lle* ‘funeral-PL-ALL’. Nevertheless, one should also keep in mind that Finnish nouns are frequently used in the nominative singular, where no case or number marking is required (in Finnish).<sup>5</sup>

<sup>5</sup> Animate direct objects may be either in the nominative or oblique, although the nominative is gaining ground especially in the case of animate direct objects. Definiteness doesn’t play a role.

Finnish Romani < Romani (in plain), Finnish (in bold)

- (10) li:ne de:velesko **armoa**  
got.3PL god.GEN mercy.PART  
“They received God’s mercy.” (Speaker 1)

The Finnish case system does not replace the Romani case, which is still in use for the Romani nouns, even though it largely replicates the functional distribution of the Finnish case system. Finnish cases are mostly restricted to Finnish nouns, although uses with Romani nouns may occur, e.g., Finnish partitive case in *paani-ta* ‘water-PART.SG’. Moreover, Finnish case is in complementary distribution with Romani adpositions, e.g., ‘in Norrköping’ occurs either with the Finnish inessive case, as in *Norrköpingi-ssä* ‘Norrköping-INESS’, or with the Romani adposition, as in *aro Norrköping*.

Case morphemes, which are “outsider” morphemes in the 4-M model, are generally known not to be inserted into the morpho-syntactic frame of another language in code-switching (Myers-Scotton 2002) but have been described in the formation of the Australian mixed language, Gurindji Kriol (Meakins 2011a, 2011b). For Finnish Romani this type of noun insertion, without integration into the Romani morphology, is facilitated by the convergence that largely affected Finnish Romani, replicating Finnish word order and case.

5.5. INTER-SPEAKER VARIATION FOR NOUNS AND VERBS. Most studies on code-switching show that there is typically considerable inter-speaker variation, including in the child speech of relatively stable bilingual communities (Backus and van der Heijden 2002). Mixed languages, on the other hand, are characterized by inter-speaker consistency with respect to patterns of mixing (O’Shannessy 2012, Meakins 2012) although as Meakins (2013) observes, there is no reason why mixed languages shouldn’t show variation as any other natural language. It has also been hypothesized that, at an intermediate stage between code-switching and mixed languages, code-switching has already become stabilized, the so-called fused lect in Auer (1998). We therefore wanted to check if inter-speaker variation affected the proportions of current contact nouns and verbs in the corpora, which is also a means of assessing the regularity of contact language content words.

In the Thrace Romani corpus, we examined four speakers who show 79-85% of Romani tokens in their overall speech and who are the best represented in the sample as far as the total number of words is concerned. Interestingly, as shown in Table 11, the female 50-year-old speaker shows 11% of Turkish nouns, whereas the three speakers who are in their 30s show 28-41% of Turkish nouns. By contrast, the distribution of Romani, Turkish, and Greek verbs is similar for all four speakers: 81-89% of Romani verbs, 6-11% of Turkish verbs, and only 1-6% of Greek verbs (Table 12).

**Table 11.** Thrace Romani corpus: Distribution of nouns per language in Thrace Romani for four speakers

	Romani		Turkish		Greek		Multiple	
	tokens	%	tokens	%	tokens	%	tokens	%
<b>S4 (F, 50)</b>	86	81	12	11	6	6	2	2
<b>S 10 (M, 34)</b>	133	63	59	28	19	9	4	0
<b>S 2 (F, 34)</b>	133	54	67	27	28	12	18	7
<b>S 9 (M, 29)</b>	106	56	78	41	4	2	2	1

The chi-squared test showed that the choice of the language for the nouns is extremely significant (p-value 8,4E-10).

**Table 12.** Thrace Romani corpus: Distribution of verbs per language in Thrace Romani for four speakers

	Romani		Turkish		Greek		Multiple	
	tokens	%	tokens	%	tokens	%	tokens	%
<b>S4 (F, 50)</b>	140	81	12	11	3	6	0	2
<b>S 10 (M, 34)</b>	247	89	29	10	3	1	0	0
<b>S 2 (F, 34)</b>	112	89	7	6	6	5	0	0
<b>S 9 (M, 29)</b>	330	89	42	11	1	0	0	0

The chi-squared test showed that the choice of the language for the nouns very significant for the verbs (p-value 5,7E-3).

The analysis of the Finnish Romani corpus also shows variability in the use of verbs and nouns by the three speakers (Table 13). For speaker 1, 21% of the nouns and 14% of the verbs are Finnish; for speaker 2, 6% of the nouns and 4% of the verbs are Finnish; and for speaker 3, 11% of the nouns and 11% of the verbs are Finnish.

**Table 13.** Finnish Romani corpus: Distribution of nouns and verbs per language in Finnish Romani for three speakers in Finnish Romani dominant speech

	Nouns				Verbs			
	Romani		Finnish		Romani		Finnish	
	tokens	%	tokens	%	tokens	%	tokens	%
<b>S1</b>	570	79	153	21	280	86	45	14
<b>S2</b>	610	94	41	6	908	96	35	4
<b>S3</b>	506	89	65	11	494	89	63	11

The chi-squared test showed that significance was extremely high for the choice of language both for the nouns and verbs (p-value 1,6E-15 and 4,2E-11 respectively).

When comparing the two corpora, we see that Finnish Romani speakers vary more in the use of Finnish nouns, whereas Thrace Romani speakers show an age-related distribution of nouns from the languages in contact. The same variability is found for

verbs in the Finnish Romani corpus, whereas the Thrace Romani speakers show similar distribution for Turkish and Romani verbs.

5.6. SUMMARY. The analysis of the Thrace Romani and Finnish Romani corpora shows a complex and atypical language mixing. The quantitative analysis of the two corpora shows that Romani is the numerically dominant language even though the Finnish Romani corpus is characterized by 20% of Finnish dominant clauses. The analysis of the length of current contact language words shows that short switches are more common in both corpora, with some longer, two or three-word switches. The analysis of the word classes shows that the mixing is not restricted to nouns but occurs for conjunctions and adverbs. Some other word classes from the contact languages, such as pronouns, are practically not present but this could also be due to the fact that all languages in contact are pro-drop languages. Moreover, the analysis of the Romani corpora shows that Turkish and Finnish verb morphology is used with Turkish and Finnish stems respectively but it is never combined with the Romani stems. This is also the case for noun morphology in the Finnish Romani corpus.

It is not clear whether this type of language mixing could be considered as an example of “composite code-switching” in Myers-Scotton’s terms, defined as the type of code-switching in which the abstract grammatical structure largely comes from one language and partially from another (Myers-Scotton and Jake 2009: 339). Convergence, a parameter which has not been dealt in detail in this paper, is problematic in that Thrace Romani has not significantly converged with Turkish and keeps for example a distinct word order (Arvaniti and Adamou 2011) but Finnish Romani has entirely converged with Finnish (Granqvist 2010). Moreover, grammatical morphemes come from the two languages in contact but are restricted to the stems of each language, in what can be thought of as the first stage prior to a possible expansion of the contact morphology to all stems (Meakins 2011a). But, if Finnish Romani can be described as an example of “composite code-switching”, we observe that the proportion of verbs and nouns from the two languages varies depending on the speakers as is often the case in classic code-switching. In contrast, Thrace Romani which is not a good candidate for the “composite code-switching” category, since it hasn’t converged with its contact language, shows little inter-speaker variation with respect to the proportion of words from each language, as posited in the definition of the “fused-lect” (Auer 1998). Without wanting to impose a new terminology, we believe that the data we present are better understood as “unevenly mixed languages,” representing a stage prior to the formation of mixed languages (McConvell and Meakins 2005, Meakins 2013).

## 6. TYPOLOGICAL AND SOCIOLINGUISTIC FACTORS

In section 5 we have presented evidence for an atypical language mixing, which could be understood as the first step in an interrupted process of mixed language creation. How is it possible that such a rare process is found in two distinct Romani communities in contact with two unrelated languages? At least two types of factors have been tackled in the literature to address the types of language mixing and the creation of mixed languages: the typological and the sociolinguistic factors.

6.1. TYPOLOGICAL FACTORS. Discussing the Thrace Romani data, Myers-Scotton (2013) suggests that some typological properties of the languages in contact may be responsible for this typologically rare verb transfer. Also, Meakins and O'Shannessy (2012) convincingly relate the intensity of the transfer of verb morphology to the typologies of the languages in contact. In the case of the Romani data though, the typology of the languages in contact doesn't seem to be the best explanatory parameter. Typologically, Romani shows mainly fusional and sometimes agglutinative patterns, whereas Turkish (Altaic) is an agglutinative language and Finnish (Uralic) shows agglutinative and some fusional patterns. Nevertheless, verb integration with contact language morphology is also observed in non-agglutinative languages: for example, North Russian Romani in contact with Russian, a synthetic language, shows a similar pattern of verb integration, see (11). Similarly to Finnish Romani and Thrace Romani, example (12) shows that variation between the Romani and the Russian verb, here "to love," is possible for a single speaker in a single sentence.

North Russian Romani (Xaladytka) < Romani (in plain), Russian (in bold)

(11) a. me **n'i mag-u** te sykl'uv-aw  
1SG NEG can-1SG COMP study-1SG  
"I can't study,"

b. man-ge čej **pamag-at'** mama-ke  
me-DAT needed help-INF mother-DAT  
"I have to help mother." (Anton Tenser, unpublished data, elicited with the Romani Morphosyntax questionnaire  
<http://romani.humanities.manchester.ac.uk/rms/>)

(12) me l'ubl'-u thud ne na kam-am parn'-in'k-a  
1SG love-1SG milk but NEG love-1SG white-DIM-PL  
"I like milk but don't like eggs." (Anton Tenser, unpublished data, elicited with the Romani Morphosyntax questionnaire  
<http://romani.humanities.manchester.ac.uk/rms/>)

Moreover, several other verb integration processes occur in a number of Romani dialects (Miklosich 1872–1880, Bakker 1997b, Elšík and Matras 2006: 324–333). For example, the use of loan verb markers originating from inflectional or derivational affixes that do not keep their grammatical value is very frequent: i.e., the forms derived from the Greek aorist *-is/-as/-os-*, as well as those derived from the Greek present tense *-iz/-az/-oz-* and *-in/-an/-on-*, which are particularly common in the Vlax Romani branch. These loan verb adaptation markers were often borrowed from Greek in Early Romani in order to accommodate the Greek loan verbs, but remained productive as loan verb markers in many Romani dialects even when Roma had lost their active knowledge of Greek (see Matras 2002 for a discussion).

6.2 SOCIOLINGUISTIC FACTORS. Specific sociolinguistic conditions appear to be at the base of this type of code-switching. Discussing the Thrace Romani mixing with Turkish, Adamou (2010) observes that it occurs in intense, high-contact settings involving tightly knit, trade-related Romani communities as opposed to settled, rural Romani communities. In a survey of several Romani varieties spoken in the Balkans, Friedman (2010) supports this analysis which allows accounting for the use of the non-integrated Turkish verbs by urban, trade-related Romani communities, and the more classic verb-integration strategy by rural Romani communities.

Moreover, similar to the process of mixed language creation, the Romani type of mixing results from conflicting attitudes between language shift and language maintenance, combining with the lack of normative pressure and general acceptance of bilingual identities (Matras 2009, O’Shannessy 2012, Meakins 2013).

But, can we evaluate the role of language shift in this mixing process? According to Friedman (2013), language shift is not relevant for the Turkish verb integration that occurs in the Romani varieties of the Balkans. Myers-Scotton (2013) on the other hand suggests that a Matrix Language Turnover is responsible for this atypical phenomenon:

That is, the explanation would be that Romani speakers were in the process of shifting to Turkish, but that this was a Matrix Language Turnover that was arrested. For socio-psychological reasons, the shift stopped. (Myers-Scotton 2013: 40).

Due to lack of data, we can only hypothesize that the type of limited language mixing we have described in this paper was created by the end of the nineteenth century by fully bilingual speakers and was transmitted during the twentieth century in both Romani communities under study. This chronology is plausible since, at least in the Balkans, a limited number of Turkish verbs with some Turkish verb morphology are still used in Romani communities which have lost contact with Turkish during the twentieth century (Igla 1996). The arrested Matrix Language Turnover analysis is plausible for the Turkish speaking communities of Greek Thrace. Intensive and

extensive contact with Turkish was frequent during Ottoman times, when Turkish was the lingua franca, the language of communication and trade in the Balkans and a shift to Turkish could have begun in some trade- and service-related Romani communities (Adamou 2010). The shift to Turkish could have been halted in the early twentieth century with the change in the sociolinguistic setting related to the end of the Ottoman Empire when, in 1923, part of Thrace was integrated into the Greek state and Turkish became a minority language. The mixed Romani variety, which the speakers call *xoraxane romane* “Turkish Romani”, was probably transmitted as such to the younger generations.

For the Romani communities of Finland, the shift to Finnish has been a long process going back to the nineteenth century when Finland was an autonomous Grand Duchy within the Russian Empire. At the end of the nineteenth century, documents indicate that Romani was restricted to specific trade-related activities and was used as a secret language, while Finnish was the everyday language (Thesleff 1899). Roma children generally did not learn Romani during childhood, but acquired the language during adolescence (Thesleff 1899). The language mixing we report on might result from a Matrix Language Turnover that may have stopped, owing to sociopolitical changes in the first part of the twentieth century. Indeed, in 1917 Finland became independent followed by other major changes in the country’s organization which took place after World War II, in the 1950s. Unlike what happened in Greek Thrace, where a new contact language, Greek, may have interrupted the shift process of Romani communities to Turkish, in Finland the parameters at play seem different. It is likely that it is the type of contact to Finnish that was modified in the early twentieth century with the rise of a monolingual model, based on a dominant state language, Finnish, which was contradictory with the ongoing mixing process. As McConvell and Meakins (2005) explain for the creation of Gurindji Kriol, specific sociopolitical changes may have a direct impact on the language attitudes of the community and radically change an existing language mixing model.

The situations in the twentieth century have evolved in different ways for the Finnish Roma in Finland and the Muslim Roma in Greece. On the one hand, for Greek Thrace Romani speakers, the shift to Turkish was interrupted in the beginning of the twentieth century. Today Turkish is used in colloquial and trade-related contexts and, along with Greek, is the language of education for the local Muslim population. In this setting, a new wave of shift to Turkish is attested among the Muslim Roma in Thrace. On the other hand, Finnish Romani speakers pursued the shifting process to their contact language during the twentieth century. Despite this gradual shift to Finnish, Romani is still used by the elders, and some young Roma and children have been reported as having mastered it (Hedman 2009). The generation of speakers who were no longer fully bilingual were probably the ones who have created a mixed language with Finnish grammar and a Romani lexicon, as an “afterlife” of Romani (Matras 2010) in a



“reversal of shift” (Meakins 2013: 182). Nowadays, some Finnish-Romani speakers master both the mixed G-L Finnish-Romani variety and the unevenly mixed Romani variety described in this paper.

## 7. CONCLUDING REMARKS

In this paper we presented the results of a quantitative analysis of free speech by Finnish Romani-Finnish speakers and Thrace Romani-Turkish-Greek speakers. The analysis of the Finnish-Romani corpus shows that the last fluent speakers of Finnish Romani use 20% of Finnish-dominant clauses and 80% of Finnish Romani-dominant clauses. Although the Finnish Romani corpus shows a majority of Romani words (65%), other sociolinguistic studies suggest that the same speakers would have used more Finnish words in an informal setting with in-group members. By contrast, the Thrace-Romani corpus, which was collected in an informal setting, shows that Romani is the dominant language of in-group interactions (80%), and the use of Turkish-dominant clauses is participant-related. Interestingly, the analysis of the dominant Finnish-Romani clauses reveals similar patterns to the Thrace-Romani speech, namely, an average of 19% of current contact language word tokens as compared to 20% in the Thrace-Romani corpus. In both corpora, verbs retain the contact language morphology in a Romani-dominant clauses (see Figure 3).

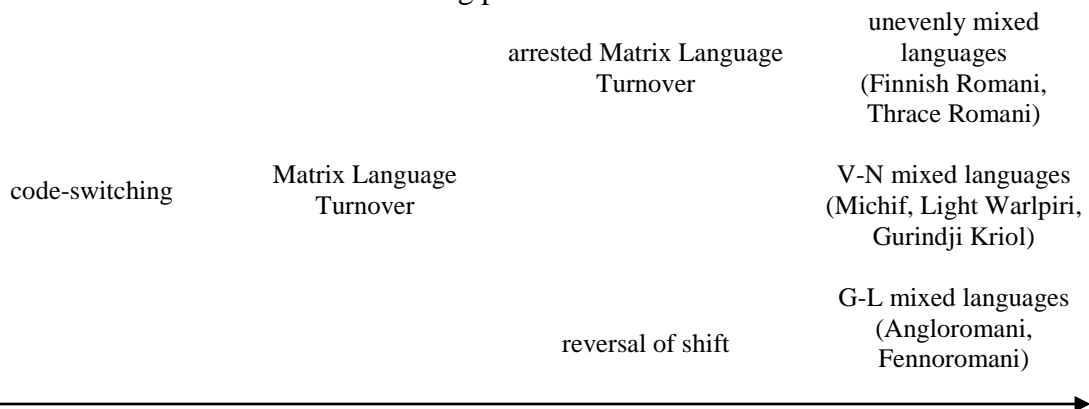
Finnish Romani-Finnish corpus		Thrace Romani-Turkish-Greek corpus			
Finnish-dominant clauses (20.6%)	Romani-dominant clauses (79.4%)	Turkish-dominant clauses (participant-related)	Romani-dominant clauses		
	Romani (81%)	Finnish (19%) V, N with Finnish morphology	Romani (80%)	Turkish (15%) V with Turkish morphology	Greek (4%)

**Figure 3.** Overview of the Finnish Romani and the Thrace Romani corpus.

We suggest here that this typologically unusual mixing is best understood as the result of an arrested Matrix Language Turnover (Myers-Scotton 2002), similar to the process of mixed language formation (see Figure 4). Nevertheless, although insertion of intact contact-language elements is a mechanism found in mixed languages, the Romani data show that the process of mixing was stopped in the communities under study. For example, the Romani current contact language verb insertions represent only 10-12% of

the total number of verbs in what resembles the early stages of code-switching before the formation of the Australian mixed languages (McConvell and Meakins 2005, O’Shannessy 2012). The Romani data thus offer evidence for two unevenly mixed languages which may be understood as the result of an interrupted process of mixed language formation.

An interesting topic to explore in future research would be the acquisition of these unevenly mixed Romani varieties by analyzing interactions between adults and adolescents and pre-adolescents in both Finland and Greece. It could also be interesting to compare the present data with quantitative data from North Russian Romani in contact with Russian, which would provide another case study for a Romani language with similar characteristics in the mixing patterns.



**Figure 4.** A path to various types of mixed languages.

## Acknowledgments

We wish to thank the three anonymous reviewers for their insightful comments and the IJB’s editor, Li Wei. Also, we are thankful to the Romani participants in Finland and in Greece. Special thanks are due to Mrs. Hellevi Hedman-Valentin for the transcription of the Finnish Romani corpus, as well as Mrs. Tuula Åkerlund, Mrs. Miranda Vuolasranta, and Mr. Juhani Pallonen who conducted the recordings. Last but not least, we thank Anton Tenser and Lameen Souag for comments on earlier versions of the paper and François Sermier for his precious help with the statistics.

## Funding

Evangelia Adamou wishes to acknowledge financial support from the French National Research Agency (ANR), project “Towards a multi-level, typological and computer assisted analysis of contact-induced language change” (ANR-09-JCJC-0121-01).

Kimmo Granqvist received funding from the Kone Foundation Language Programme, project “Finnish Romani and other Northern dialects of Romani in the Baltic Sea area” (2013–2016).

## Abbreviations

Glosses follow the Leipzig glossing rules in the augmented version of the Corpus for Afroasiatic language program <http://corpafroas.tge-adonis.fr/glosses.html>.

## References

- Adamou, Evangelia. 2010. Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish, *Language in society* 39 (2): 147–171.
- Adamou, Evangelia. 2012. Verb morphologies in contact: evidence from the Balkan area. In Martine Vanhove, Thomas Stolz, Hitomi Otsuka and Aina Urdze (eds.), *Morphologies in contact*, 143–162. Berlin: Akademie Verlag.
- Arvaniti, Amalia and Evangelia Adamou. 2011. Focus Expression in Romani. In M. Byram Washburn et al. (eds). *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 240–248. Somerville, MA: Cascadia Proceedings Project.
- Auer, Peter. 1998. From code-switching via language mixing to fused lects: toward a dynamic typology of bilingual speech. *Interaction and Linguistic Structures* 6: 1–28.
- Backus, Ad and Hanneke Van der Heijden. 2002. Language mixing by young Turkish children. *Psychology of Language and Communication* 6 (1): 55–73.
- Bakker, Peter. 1995. Genesis of Caló and other Iberian Para-Romani varieties. In Yaron Matras (ed.), *Romani in Contact*, 125–150. Amsterdam: John Benjamins.
- Bakker, Peter. 1997a. *A language of our own. The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Oxford: Oxford University Press.
- Bakker, Peter. 1997b. Athematic morphology in Romani: the borrowing of a borrowing pattern. In Yaron Matras, Peter Bakker and Hristo Kyuchukov (eds.), *The typology and dialectology of Romani*, 1–21. Amsterdam: Benjamins.
- Bakker, Peter and Maarten Mous. 1994. Introduction. In Peter Bakker and Maarten Mous (eds.), *Mixed languages: 15 case studies in language intertwining*, 1–11. Amsterdam: IFOTT.
- Bakker, Peter and Yaron Matras. 2013. Introduction. In Bakker Peter and Yaron Matras (eds), *Contact languages*, 1–14. Boston and Berlin: Walter de Gruyter.
- Borin, Lars. 2000. A corpus of written Finnish Romani texts. In Donncha O. Croinin (ed.), *LREC 2000. Workshop proceedings. Developing language resources for minority languages: reusability and strategic priorities*, 75–82. Athens: ELRA.

- Carling, Gerd, Lenny Lindell and Gilbert Ambrazaitis. in press. *Scandoromani. History of a mixed language*. Leiden: Brill.
- Cortiade, Marcel. 1991. Romani versus Para-Romani. In Bakker, Peter and Marcel Cortiade (eds.), *In the Margin of Romani. Gypsy languages in contact*, 1–15. Amsterdam: Institute for General Linguistics.
- Friedman, Victor. 2000. Romani in the Balkan Linguistic League. In Christos Tzitzilis and Charalambos Symeonides (eds.), *Valkanike Glossologia: Sygkhronia kai Diakhronia/Balkanlinguistik: Synchronie und Diachronie*, 95-105. Thessaloniki: University of Thessaloniki.
- Granqvist, Kimmo. 2007. *Suomen romanin äänne- ja muotorakenne*. Suomen Itämaisen Seuran Suomenkielisiä julkaisuja 36. Kotimaisten kielten tutkimuskeskuksen julkaisuja 145. Helsinki: Yliopistopaino.
- Granqvist, Kimmo. 2010. Lost features in Finnish Romani. Paper presented at 9th International Conference on Romani Linguistics, Helsinki, September 2–4, 2010.
- Elšík, Victor and Yaron Matras. 2006. *Markedness and language change*. Berlin and New York: Mouton de Gruyter.
- Friedman, Victor. 2010. Turkish grammar in Balkan Romani: Hierarchies of markedness in Balkan linguistics. *Balkanistica* 23: 107-124.
- Friedman, Victor. 2013. Compartmentalized Grammar: The Variable (Non)-Integration of Turkish Verbal Conjugation in Romani Dialects. *Romani Studies* 23(1): 107-120.
- Hedman, Henry. 1996. *Sar me sikjavaa romanes. Romanikielen kielioppiopas*. Jyväskylä: Opetushallitus.
- Hedman, Henry. 2009. *Suomen romanikieli: sen asema yhteisössään, käyttö ja romanien kieliaseenteet*. Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 8. Kotimaisten kielten tutkimuskeskus.
- Igla, Birgit. 1996. *Das Romani von Ajia Varvara – Deskriptive und historisch vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.
- Krauss, Michael. 2006. Classification and terminology for degrees of languages endangerment. In Matthias Brenzinger (ed.), *Language Diversity Endangered*, 1-8. Berlin: Mouton de Gruyter.
- Koivisto, Viljo. 1987. *Rakkavaha romanes. Kaalengo tšimbako sikjibosko liin*. Helsinki: Ammattikasvatustieteiden tutkimuskeskus and Valtion painatuskeskus.
- Koivisto, Viljo. 1994. *Romano–finitiko–angliko laavesko liin. Roman–suomi–englanti sanakirja. Romany-Finnish-English Dictionary*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 74. Helsinki: Valtion painatuskeskus.
- Kopsa-Schön, Tuula. 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.

- Kovanen, Petra 2010. Koodinvaihtelu romanikielisessä keskustelussa. Unpublished Master's thesis. University of Helsinki.
- Matras, Yaron. 2002. *Romani: a linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2003. Mixed languages: Re-examining the structural prototype. In Matras and Bakker (eds), *The mixed language debate*, 151–76. Berlin: Mouton de Gruyter.
- Matras, Yaron. 2005. The classification of Romani dialects. A geographic-historical perspective. In Dieter Halwachs and Barbara Schrammel (eds.), *General and applied Romani linguistics*, 7–26. Munich: Lincom Europa.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2010. *Romani in Britain: The afterlife of a language*. Edinburgh: Edinburgh University Press.
- McConvell, Patrick and Felicity Meakins. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25 (1): 9-30.
- Meakins, Felicity. 2011a. Borrowing contextual inflection: evidence from northern Australia. *Morphology* 21: 57-87.
- Meakins, Felicity. 2011b. *Case-marking in contact: The development and function of case morphology in Gurindji Kriol*. Amsterdam: Benjamins.
- Meakins, Felicity. 2012. Which Mix? - Codeswitching or a mixed language? – Gurindji Kriol. *Journal of Pidgin and Creole Languages* 27 (1): 105-140.
- Meakins, Felicity. 2013. Mixed Languages. In Bakker Peter and Yaron Matras (eds), *Contact languages*, 159-228. Boston and Berlin: Walter de Gruyter.
- Meakins, Felicity and Carmel O'Shannessy. 2012. Typological constraints on verb integration in two Australian mixed languages. *Journal of language contact* 5 (2): 216-246.
- Miklosich, Franc. 1872–1880. *Über die Mundarten und Wanderungen der Zigeuner Europas*. vol. 3. Vienna: Karl Gerold's Sohn.
- Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-mixing*. Cambridge University Press: Cambridge.
- Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Code-switching*. Oxford: Clarendon press.
- Myers-Scotton, Carol. 2002. *Contact linguistics, bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, Carol and Janice Jake. 2009. A universal model of code-switching and bilingual processing and production. In Barbara Bullock and Almeida Jacqueline Toribio (eds.), *Linguistic code-switching*, 336-357. Cambridge: Cambridge University Press.

- Myers-Scotton, Carol. 2013. Paying attention to morpheme types: making borrowability more precise. In Carole de Feral (ed.), *In and out of Africa languages in question*, 31-42. Louvain: Peeters.
- O'Shanessy, Carmel. 2012. The role of codeswitched input to children in the origin of a new mixed language. *Linguistics* 50 (2): 305-340.
- O'Shanessy, Carmel. 2013. The role of multiple sources in the formation of an innovative auxiliary category in Light Warlpiri, a new Australian mixed language. *Language* 89 (2): 328-353.
- Poplack, Shana and Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24 (3): 279-315.
- Rusakov, Alexander. 2001. The North Russian Romani dialect: interference and code switching, In Östen Dahl and Maria Koptjevskaja-Tamm (eds), *Circum-Baltic languages*, 313–338. Amsterdam and Philadelphia: Benjamins.
- Thesleff, Arthur 1899: *Finlands zigenare. En etnografisk studie*. Helsinki.
- Treffers-Daller, Jeanine. 1994. Mixing two languages: French-Dutch contact in a comparative perspective. Berlin: Mouton de Gruyter.
- Valtonen, Pertti. 1968. *Suomen mustalaiskielen kehitys eri aikoina tehtyjen muistiinpanojen valossa*. Licentiate thesis. Helsinki.
- Vuorela, Katri and Lars Borin. 1998. Finnish Romani. In Ailbhe Ó Corráin and Séamus Mac Mathuna (eds.), *Minority Languages in Scandinavia, Britain and Ireland* 3, 51–76. Uppsala.

### Author biographies

Evangelia Adamou is Associate Researcher at the National Centre for Scientific Research, France. Her main research interests are language contact, syntax, and information structure.

Kimmo Granqvist is Adjunct Professor, University Lecturer at the University of Helsinki, Finland. His main research interests are language contact, linguistic variation and change, phonology and morphosyntax.