



**HAL**  
open science

# How sentence processing sheds light on mixed language creation

Evangelia Adamou

► **To cite this version:**

Evangelia Adamou. How sentence processing sheds light on mixed language creation. Eeva Sippola; Maria Mazzoli. *New Perspectives on Mixed Languages. From Core to Fringe*, De Gruyter, In press, 978-1-5015-1125-7. halshs-02442565

**HAL Id: halshs-02442565**

**<https://shs.hal.science/halshs-02442565>**

Submitted on 22 Jan 2020

**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.

## How sentence processing sheds light on mixed language creation

Evangelia Adamou

CNRS (French National Centre for Scientific Research)

### Abstract

Auer (1999) and O’Shannessy (2012) suggest that codeswitching can become conventionalized, i.e., what is dubbed a “fused lect”, and eventually give rise to a “mixed language”. In this paper I propose that the study of sentence processing can shed light on this process. I discuss recent experimental data from a typologically rare form of mixing, variably termed “fused lect” (Adamou 2010) and “unevenly mixed language” (Adamou and Granqvist 2015), characterized by the conventionalized use of Turkish verbs together with Turkish morphology in a Romani environment. Specifically, Adamou and Shen (2019) conducted two on-line experiments, a picture choice task with sentence auditory stimuli (37 participants) and a word recognition task in sentence context (49 participants). Results from these experiments indicate that language switching costs depend on the degree of conventionalization and support usage-based approaches to language processing. I argue that these findings also lend support to the categorization of fused lects as an intermediate form between codeswitching and mixed languages.

Keywords: fused lect; mixed languages; codeswitching; sentence processing; language switching costs

### 1. Introduction

Linguists are increasingly paying attention on the ways bilinguals may alternate between two languages in one situation and with a single interlocutor, and how, under some specific sociolinguistic circumstances, systematic codeswitching may lead to the genesis of a novel, mixed language. Furthering the naturalistic evidence on mixed languages, I suggest that the study of sentence processing and of language switching costs can shed light on the creation of mixed languages. I illustrate this proposal through results from sentence processing of a so-

called “fused lect” (Adamou 2010) or “unevenly mixed language” (Adamou and Granqvist 2015), comprising a majority of Romani linguistic material combined with the conventionalized use of Turkish elements, including verbs and verb morphology (Adamou and Shen 2019). Though in the study by Adamou and Shen (2019) the focus was on the importance of the Romani data for the discussion of language switching costs, in the present paper, I discuss these results with respect to the process of formation of mixed languages. In the remainder of this introduction, I present a brief overview of the literature on mixed languages (in 1.1.) and language switching costs (in 1.2.). In 1.3. I argue for the advantages of combining the two research paradigms with benefits for both fields: on the one hand, allowing to deepen our understanding of mixed languages, on the other hand, to understand more fully human cognition and the effects of cultural and linguistic variation.

### 1.1. Mixed languages

In past decades significant progress has been made in documenting “mixed languages” and understanding their formation (see among others Thomason and Kaufman 1988; Bakker and Mous 1994; Thomason 1995; Bakker 1997; Muysken 1997; Auer 1999; Myers-Scotton 1998, 2002; Matras 2000, 2003). Mixed languages can be broadly defined as “the result of the fusion of two identifiable source languages, normally in situations of community bilingualism” (Meakins 2013: 159).

At present, two well-established structural types of mixed languages have been identified (Meakins 2013):

- a) G(rammar)–L(exicon) mixed languages, which draw the grammar from one language and the lexicon from another, for example, Ma’á, a Bantu language spoken in Tanzania with elements from Cushitic languages (Bakker and Mous 1994).
- b) V(erb)–N(oun) mixed languages, which exhibit structural mixing, with nouns from one language and verbs from another. A good example of V-N mixed languages is Michif, a Cree-French mixed language spoken by the Métis (Bakker 1997). Michif combines the verb system from Cree and the noun system from French. Other well-known examples of V-N languages are the two Australian mixed languages, Gurindji Kriol (Meakins 2012) and Light Warlpiri (O’Shannessy 2013).

The process of mixed language genesis has been the locus of heated debate in the early studies. Some researchers considered mixed languages as the product of extraordinary mixing processes, for example, of language intertwining (Bakker and Mous 1994), or lexical re-orientation when the ancestral language anchors the predication (i.e., provides the inflection of the finite verbs) and selective replication when the socially dominant language anchors the predication (Matras 2000, 2003), eventually going through transitional stages in the anchoring of the predication as observed at the level of the discourse (Matras et al. 2007). Other researchers suggested instead that mixed languages emerge following the conventionalization of codeswitching patterns (Auer 1999; Myers-Scotton 2002). According to Myers-Scotton, intrasentential codeswitching, that is, codeswitching at the level of the clause, comprises any number of words from the Matrix Language, the language that sets the grammatical frame, and at least one word from the Embedded Language, the language that mainly contributes content morphemes. Myers-Scotton (1998) proposed that mixed languages are created when there is a “Matrix Language Turnover” which is interrupted; this is dubbed “Arrested Matrix Language Turnover”. Moreover, Myers-Scotton (2002: 105, 269) tentatively suggested that speakers may go through a stage of “composite codeswitching” during which, within a single clause, they draw the Matrix Language from more than one source. Auer (1999), in turn, hypothesized that “codeswitching”, that is, the meaningful alternation between two languages that can index aspects of the situation or of the speaker, becomes “codemixing”<sup>1</sup> when generalized in a bilingual community, and then eventually gives rise to a “fused lect” followed by an independent “mixed language”. Auer’s fused-lects are defined as follows: “the use of one ‘language’ or the other for certain constituents is obligatory in FLs [fused lects]; it is part of their grammar, and speakers have no choice” (Auer 1999: 321). Although Auer cites evidence from the Sinti dialect of Romani to illustrate fused lects, Adamou (2010) argued that Greek Thrace Romani data offer a more adequate example for this stage as the Turkish verbs are neither clearly borrowings nor codeswitching insertions. Finally, mixed languages differ from fused lects as they comprise innovative grammatical patterns, which did not exist in the source languages (for evidence of this last stage see Meakins 2013; O’Shannessy 2013). The continuum hypothesis is illustrated in Figure 1, adapted from Auer with the difference that I visualize mixed languages as a different stage from fused lects.

---

<sup>1</sup> Note that a variety of definitions of code-mixing can be found in the literature.

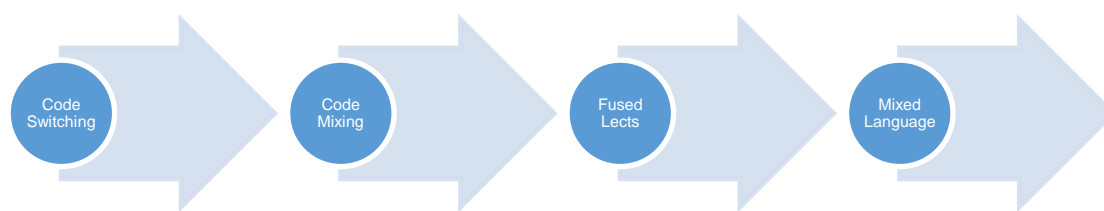


Figure 1. A continuum from codeswitching to fused lects and mixed languages, adapted from Auer 1999

Recently, the analysis of the language use of older and younger cohorts in the community provided decisive evidence in support of the view that mixed languages result from the intensification of general contact phenomena such as codeswitching (McConvell and Meakins 2005; Meakins 2012; O’Shannessy 2012). More evidence in support of intermediate stages preceding the creation of independent mixed languages come from two distantly related Romani varieties spoken in Greece and in Finland. In these Romani varieties, morphologically intact verbs from the languages they are in contact with, Turkish and Finnish respectively, are inserted into a Romani-dominant speech by all proficient Romani speakers (Adamou 2010; Adamou and Granqvist 2015).

From a sociolinguist perspective, the creation of mixed languages is favored in highly bilingual communities, in particular, when there are conflicting processes of language shift and language maintenance (Thomason and Kaufman 1988; O’Shannessy 2012; Meakins 2013). McConvell and Meakins (2005) describe how Gurindji Kriol was formed under very specific sociopolitical changes that radically modified the community’s language attitudes and the existing language mixing model. Bakker (1994: 24) noted that mixed languages can also be created in mixed households, as is the case of Michif.

Although some mixed languages arise from the conventionalization of codeswitching, they exhibit characteristics that greatly differ from classic codeswitching. This can be illustrated with an example from a V-N mixed language from Australia, Light Warlpiri. In Light Warlpiri most verbs come from English/Kriol, verb structure comes from both Warlpiri

and English/Kriol, and nominal structure from Warlpiri (O’Shannessy 2013).<sup>2</sup> See example (1).

Light Warlpiri (Australia) < Warlpiri (in plain) and English/Kriol (in bold)

- (1) *Junga mayi nyuntu **yu-m** go wati-kari-kirl mayi?*  
 true Q 2SG 2SG.S-NFUT go man-other-COM Q  
 ‘Is it true that you went with another man?’  
 (O’Shannessy 2013: 330)

Insertions of verbs with verb morphology in a clause otherwise consisting of lexical and grammatical elements from the other language depart from classic forms of codeswitching where isolated words are generally morphologically integrated (Poplack and Dion 2012). Specifically, the use of non-integrated verbs from a given language inserted in the speech of another language is cross-linguistically rare (Wohlgemuth 2009). Myers-Scotton and Jake (2014, 2017) suggest that bilinguals either prefer non-finite verbs from the Embedded Language or morphologically integrated Embedded Language-verbs with the inflection of the Matrix Language. The 4-M model,<sup>3</sup> in particular, predicts that agreement in codeswitching will always come from the Matrix Language (Myers-Scotton 2002). According to Myers-

<sup>2</sup> In addition, Light Warlpiri exhibits a new auxiliary that draws on Warlpiri, English, and Kriol. Also Kriol verbs select ergative case marking from Warlpiri.

<sup>3</sup> The 4-M model distinguishes four types of morphemes: “content morphemes”, “early system morphemes” (e.g., determiners, derivational prepositions, particles in phrasal verbs, derivational and plural markers in noun phrases, some tense and aspect markers in verbal clauses, subordinating and coordinating conjunctions), and two types of “late system morphemes”: “bridges” (e.g., elements that join together two NPs and complementizers that join together two clauses) and “outsiders” (e.g., agreement morphemes and some case markers). In terms of the speech production model, there is a distinction between the “Mental Lexicon” (i.e., lemmas underlying morphemes), the “Lexical-Conceptual” level (i.e., where content morphemes and early system morphemes are elected), and the “Formulator” (i.e., the production mechanism that puts together the larger constituents that indicate the structure of the clause). While content morphemes and early system morphemes are salient in the mental lexicon, late system morphemes are not salient until the level of the Formulator.

Scotton and Jake (2017), the reason why Embedded Language-verbs do not keep Embedded Language-morphology in codeswitching could be due to the fact that their production would be “more costly” as speakers would need to check agreement in addition to checking at the lexical-conceptual level. Although the authors do not make any predictions in terms of processing costs, this is an interesting hypothesis that can be discussed in the light of the results reported in the study by Adamou and Shen (2019). In the following section I present an overview of previous studies on language switching costs.

## 1.2. Language switching costs

Among research questions that remain open is whether the simultaneous use of two languages is associated to cognitive costs. Using experimental methods, several studies indicate high costs, whether in production or in comprehension (see among others Soares and Grosjean 1984; Grainger and Beauvillain 1987; Meuter and Allport 1999; Thomas and Allport 2000; Jackson et al. 2001; Alvarez et al. 2003; Costa and Santesteban 2004; Proverbio et al. 2004). However, recent studies increasingly show that there are either low or no costs associated with “language” switching (see among others Moreno et al. 2002; Jackson et al. 2004; Ibáñez et al. 2010; Gullifer et al. 2013; Mosca and Clahsen 2016; Adamou and Shen 2019; Johns, Valdés Kroff, and Dussias 2019).

To understand the conflicting evidence, one needs to distinguish between simultaneous use of two languages when the speaker/comprehender is translating (Ibáñez et al. 2010) from simultaneous use when the speaker/comprehender is forced to switch between languages artificially, from simultaneous use when she is codeswitching (Chan et al. 1983; Gullifer et al. 2013). For example, professional translators (who need to comprehend sentences in one language and produce them into another language) or frequent codeswitchers (who need to comprehend sentences in one or more languages and produce a response using sentences from that same pair of languages) will not exhibit the same costs as speakers who do not have the habit of translating or codeswitching.

In addition, codeswitching stimuli need to be as close as possible to the communication habits of the codeswitcher as otherwise unexpectedness effects may arise (Moreno et al. 2002). For example, it is important to use auditory stimuli when codeswitching is generally performed in oral. Moreover, extensive research on codeswitching has shown that bilinguals tend to conform to specific patterns of codeswitching and that, despite variability, not all possible combinations of elements from two or more languages are attested among

codeswitchers. It is therefore equally important that the experimental stimuli respect switch sites, length of switch, and constraints such as the use of verbal inflectional morphology from one language or another, etc. In the past, the stimuli used in many studies on sentence processing were artificial, for example, single switched words were inserted into longer sentences along with their native morphology, something that, as the previous section showed, is rather rare in classic codeswitching that the participants of these studies are accustomed to (e.g., for nouns in Moreno et al. 2002; Proverbio et al. 2004; for nouns and verbs in Hatzidaki et al. 2011).

The importance of taking into consideration the communicational habits of the speakers/comprehenders in the study of the bilingual brain is expressed by Abutalebi and Green (2016). In Green and Abutalebi (2013) the authors elaborate the “adaptive control hypothesis” that assumes that the language control network is flexible enough to adapt to the needs of the interactional setting. Specifically, Green (1998, 2011) claims that bilinguals who codeswitch frequently in their everyday lives rely on the joint activation of the two languages, whereas bilinguals who do not frequently codeswitch rely on their language control network. More recently, Adamou and Shen (2019) and Johns, Valdés Kroff and Dussias (2019) showed that language switching costs align with the general codeswitching habits of the members of a speech community as documented in corpus studies. Though speech communities are idealizations that serve as analytical tools, there is ample empirical evidence showing that children tend to acquire linguistic patterns similar to those that they were exposed to, and that adult speakers tend to converge with the productions of the individuals they interact with, whether in the long run or temporarily during an interaction.

### 1.3. Combining the two research paradigms (i.e., mixed languages and language switching costs)

At present, there is no research on processing among speakers of mixed languages. Indeed, research on mixed languages has primarily relied on the analysis of ecologically valid data that allow for the examination of the constraints and social significance of the mixed languages. In contrast, in experimental approaches to bilingual sentence processing researchers rely on controlled data produced in a laboratory environment with natural corpus data sometimes being used as basic frequency data.

This lack of processing studies among speakers of mixed languages is also due to the difficulties of conducting experimental research outside the laboratory as well as the need to



adapt the research paradigms to groups who are not accustomed to this sort of task, and certainly not in a language that is not tied to formal education. Indeed, studies on the effects of literacy on oral language processing show consistent differences between literate and illiterate subjects (see Tarone and Bigelow 2005 for an overview), and yet research on bilinguals who are illiterate in both languages is virtually non-existent. An encouraging sign is that researchers are now starting to adapt successfully experimental and, in particular, psycholinguistic methods in the field (see O’Shannessy and Meakins 2012; Lipski 2016; Adamou and Shen 2019). However, I would like to stress here that these experimental studies were conducted following a first stage of extensive participant observation that allowed the researchers to discover the research questions that were relevant for the languages under study and not merely transpose research questions from well-studied languages.

I believe that taking into consideration lesser-documented forms of language mixing from a greater variety of groups can contribute to our understanding of bilingual processing at several levels. At a descriptive level, studying language processing among bilingual speakers of conventionalized codeswitching and mixed languages will allow us to better understand how these contact varieties are processed. At a theoretical level, widening the typological database by considering rare forms of codeswitching and language mixing may shed some light on the reasons why such language contact outcomes are cross-linguistically rare by disentangling the cognitive and the sociolinguistics factors. Moreover, for research on bilingual cognition to reflect cultural and linguistic diversity, it should move beyond the populations that are currently being studied extensively, the so-called “Western, educated, industrialized, rich, and democratic” populations or WEIRD (see Henrich, Heine, and Norenzayan 2010 for the use of this acronym to refer to the exceptional nature of such samples and call for more research in the field of human psychology among “non-WEIRD” populations).

To illustrate some of the contributions of this approach, I focus on the first experimental evidence that come from a typologically rare form of codeswitching as spoken by Romani-Turkish simultaneous bilinguals residing in the Thrace region of Greece (Adamou and Shen 2019).

## 2. The Romani data in relation to the discussion on mixed language creation

### 2.1. Background on the mixed Romani-Turkish variety

Several corpus-based studies have shown that Muslim Roma from Greek Thrace are using Turkish verb morphology with Turkish verbs inserted into Romani dominant speech (Adamou 2015; Adamou and Granqvist 2015; Adamou 2016). The resulting mixed variety is referred to by its speakers as *Xoraxane Romane* ‘Turkish Romani’ and no negative values are associated with it. Part of the Romani-Turkish recordings were quantitatively analyzed (Adamou and Granqvist 2015). The resulting corpus, of approximately 6,000 word tokens, comprises data from story-telling and interviews with the researcher as well as in-group conversations between 21 Roma speakers recorded in the community. The corpus was transcribed, aligned to the sound, annotated for sentences, words, morphemes, and parts of speech. Words and morphemes were tagged for language as a heuristic tool to discover the precise ways in which the two languages are combined while remaining agnostic about whether individuals processed them as alternations of two languages or as a single one, a question that was later tackled in Adamou and Shen (2019). More specifically, words and morphemes were tagged as “Romani” when the annotator could identify them as being of Indic origin or borrowed from past-contact languages. By contrast, words from the two current-contact languages that the speakers use in their everyday lives, were tagged as either “Turkish” or “Modern Greek”. A fourth tag, “multiple”, was applied to words for which more than one languages of origin were possible.

Analysis of this corpus reveals a numerically-dominant language, Romani, as speakers use on average roughly 15% Turkish words from all parts of speech with the exception of free pronouns. Adamou (2016) and Bullock et al. (2018) observe that the ratio of all tokens is a good proxy to determine the Matrix Language of the majority of sentences in a corpus. However, they also note that these ratios cannot predict the language that determines word order or the language of morphology. Indeed, in the Romani-Turkish corpus, though word order is consistently Romani, there is a split in morphology: while Turkish nouns inflect in Romani for case, number, and gender, like Romani nouns do, Turkish verbs systematically combine with the Turkish person, tense-aspect-modality morphology, and valency morphemes, as opposed to Romani verbs that combine with Romani verbal morphology (Adamou and Granqvist 2015). Adamou and Arvaniti (2014: 228) further note that speakers adopt mixed strategies of phonological adaptation for the Turkish lexical material when

speaking in Romani-Turkish.<sup>4</sup> An example illustrating the use of Turkish verbs in Romani-Turkish is provided in (2).

Greek Thrace Romani: Turkish in bold

(2)	<i>ep</i>	<i>me</i>	<i>ka</i>	<i>dikh-av</i>	<i>kale</i>
	always	1SG.NOM	will	look-1SG	them
	<i>me</i>	<b><i>da</i></b>	<b><i>səndəm</i></b>		
	1SG.NOM	FOC	tired.PRET.1SG		
	<i>me</i>	<b><i>da</i></b>	<i>mang-av</i>	<b><i>dineneəm</i></b>	
	1SG.NOM	FOC	want-1SG	rest.OPT.1SG	

‘Am I **always** the one to look after them? I’m **tired** of it! Me **too**, I want to **rest**.’  
(Adamou and Shen 2019: 3)

Recall that according to the Matrix Language Frame model, the Matrix Language is the language that supplies the system morphemes, in particular the outsiders in the subsequent 4-M model, namely agreement verb morphology and some case morphemes as they establish the relationship between an argument and a verb. In example (2) one has to acknowledge the use of two Matrix Languages as there is agreement between the subject (Romani pronoun with Romani nominative case) and the finite verb ‘tired’ (in Turkish). In addition, there is agreement between the subject (Romani pronoun with Romani nominative case) and the Romani deontic finite verb ‘want’ (with present tense and indicative mood) as well as its complement clause verb, the Turkish finite verb ‘rest’ (with optative mood). In Romani-Turkish, the use of agreement verb morphology from two languages, Romani and Turkish, and case from only one of them, Romani, can be observed not only within a single clause, but overall throughout the spontaneous interactions. The opposite pattern, however, is not

---

<sup>4</sup> For example, Romani speakers use several Turkish vowels that are not part of Romani phonology (e.g., [y u œ]), and yet they do not use Turkish /h/ in either Ottoman-period borrowings (e.g., *maala* ‘neighborhood’ from Turkish *mahal*), or in more recent borrowings, (e.g., *apo* ‘pill’ from Turkish *hap*). Also, they do not use vowel harmony when the nouns are integrated into Romani morphology; for Turkish clusters, including in the verbs, metathesis is frequent (e.g., Turkish *anlamayacak* ‘he/she will not understand’, becomes [al'nama,dzak]).

attested: for example, clauses comprising Turkish pronouns with Romani finite verbs are not encountered. This points to the existence of specific, conventionalized patterns in the way the two languages can be brought together in a clause or in the discourse.

The combination of Turkish verbs with Turkish verb morphology and Romani and Turkish nouns with Romani morphology is reminiscent of so-called V-N mixed languages where the verbs come from one language and the nouns from another. And yet, Adamou and Granqvist (2015) observe that the mixing process is not complete as only a small ratio of verbs come from Turkish (i.e., 12%), while the majority of verbs come from Romani and combine with Romani verb morphology. Although comparable frequency data are not available, full-fledged mixed languages such as Light Warlpiri and Gurindji Kriol are said to “derive their lexicon relatively evenly from their source languages” (O’Shannessy and Meakins 2012: 384). Adamou and Granqvist (2015) therefore suggest that the Romani data are representative of the early stages of mixed languages, as documented for the Australian languages (McConvell and Meakins 2005; O’Shannessy 2012). Unlike what happened in the Australian languages, the mixing process has been interrupted in Romani.

Indeed, Myers-Scotton (2013) suggests the following scenario discussing the Romani-Turkish data from Adamou (2010): “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). Adamou and Granqvist (2015) agree that an Arrested Matrix Language Turnover could account for the Romani-Turkish variety. They hypothesize that a shift to Turkish in the late 19<sup>th</sup> century may have been interrupted in the early 20<sup>th</sup> century when the Ottoman Empire collapsed. At that time, new borders modified the traditional mobility of Roma populations who were involved in trade throughout the Ottoman Empire. The area of Thrace eventually became part of the Greek State in the 1920s. This change gave rise to an entirely different sociopolitical setting where Turkish became the minority language and Greek the language of the State. A strong argument in favor of the formation of the Romani-Turkish variety in the early 20<sup>th</sup> century comes from the fact that similar patterns of Romani-Turkish mixing are documented in other contemporary Romani-speaking communities in the Balkans in which Turkish is no longer spoken for the past hundred years. The fact that Turkish is still spoken nowadays in Greek Thrace is therefore not what gives rise to the Romani-Turkish variety that we observe, but merely allows the mixing to be further developed with new lexical additions and innovations.

More generally, similar mixing processes as the one documented in Greek Thrace can also be found in a number of Romani dialects in contact with a variety of languages. For example, Crimean Romani (Elšík and Matras 2006: 135) and several other Romani dialects spoken in the Balkans (Friedman 2013) use Turkish verbs similarly to Thrace Romani. Finnish Romani also draws verbs from Finnish together with Finnish morphology (Adamou and Granqvist 2015). North Russian Romani (Rusakov 2001), Soviet Romani, and Lithuanian Romani (Elšík and Matras 2006: 135) all draw verbs from Russian and use them together with Russian morphology. See an example from the written corpus of Soviet Romani, in (3), demonstrating that the process is highly conventionalized as it is also put into writing.

Soviet Romani: Russian in bold

(3)

<i>Amə</i>	<b><i>organizuj-em</i></b>	<i>bar-ə</i>	<i>sovetsk-a</i>	<i>xulaib-əna</i>
we.NOM	organise-PRS.1PL	big-NOM.PL	Soviet-NOM.PL	enterprise-NOM.PL

We organize large Soviet enterprises.’

(Nevo Drom, No. 1, 1931; excerpt from a million-word corpus of Soviet Romani texts from the 1920s-1930s; p.c. by Kirill Kozhanov)

The use of these rare mixing processes in a variety of Romani dialects that have no contact with one another suggests that sociolinguistic factors may be at play, shared by all the above-mentioned Romani communities. For example, these tightly knit, trade-related Romani communities exhibit extensive and intensive bilingualism, as opposed to other Romani rural communities with different communication habits (Adamou 2010).

We now turn to present a brief overview of the sociolinguistic context in the Romani community of Drosero in Greek Thrace and the rich linguistic repertoires that can be encountered (for a first account see Adamou 2010). Nowadays, the Romani community under study is settled at the outskirts of the town of Xanthi, in Drosero; see map in Figure 1. The community is large, with more than 5,000 members. Most Roma from Drosero have low

socioeconomic status, with high rates of unemployment. People generally work in small trades or as seasonal workers while men are also working in ships.<sup>5</sup>

The mixed Romani-Turkish variety is the home language and the language of in-group communication. Most children receive input in this Romani-Turkish variety, in the (extended) family, in the community of Drosero as well as in other Muslim Romani communities in Greek Thrace, for example, in the Romani community settled in the outskirts of the town of Komotini as intermarriages and exchanges between members of these communities are frequent. Roma from Drosero also use Romani with other Romani-speaking communities in Greek Thrace (such as Christian Roma merchants, *Dasikane Roma*), and beyond (in other Greek cities or in the Balkans). The Romani variety they use in these interactions is not the Romani-Turkish variety, *Xoraxane Romane*, in that they are careful to avoid the Turkish lexicon, and in particular the Turkish verbs. Indeed, metalinguistic commentaries indicate that speakers of Romani-Turkish may be conscious of the Turkish/non-Turkish origin of the lexicon they use as they also speak Turkish. For example, Adamou and Arvaniti (2014: 226) note that their consultants recognized 'kavako' 'tree' as a borrowing from Turkish and considered more appropriate for the recordings the use of the 'Romani' form *ko'patfi*, even though they were not aware that the latter was of Romanian origin as Romanian is not currently spoken in the community.

---

<sup>5</sup> See *Little Home* (<https://vimeo.com/294901471>), a short film resulting from participatory digital story-telling among Roma youth in Drosero who film and interview several members of the community about their profession. In a second short film, *Aver Than* 'Other Place' (<https://vimeo.com/294780272>), teenagers talk (in *Xoraxane Romane*) about how they view life in their community. The production was possible with funding from the French National Research Agency (ANR) for the programme Empirical Foundations of Linguistics (ANR-10-LABX-0083), in collaboration with the Greek NGO Caravan Project and the local Romani NGO Hope.

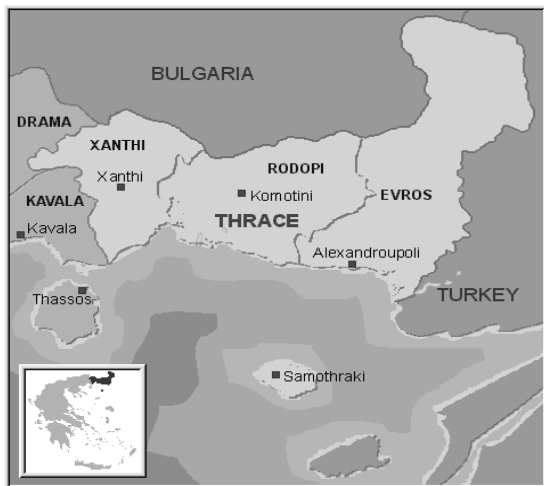


Figure 1. Map of Thrace, Greece. The study was conducted in the town of Xanthi.

In addition, most Roma children from Drosero also receive input by their families and community members in Turkish. Even though this Turkish variety has not been the object of extensive study, it can be tentatively described as a combination of the local Balkan Turkish variety, as traditionally spoken by other Muslim communities in Greek Thrace, and the varieties spoken in Turkey, as trade and visits to nearby Turkey are increasingly frequent. A distinction, however, needs to be made between Roma who have had simultaneous exposure to Turkish and Romani-Turkish (i.e., before the age of 3), those who had sequential exposure to Turkish (i.e., after the age of 3), and those who had early exposure to Turkish with little to no exposure to Romani-Turkish. Indeed, at present, a wave of shift to Turkish is ongoing in the community and is completed among several Romani families. To understand such a language shift from the traditional minority language, Romani(-Turkish), to another minority language, Turkish, and not the official language of the Greek State, Greek, one needs to keep in mind that Turkish has become in recent years an important language for the identity of Muslim Roma as it is strongly related to the Muslim religion. This is supported by the fact that Turkish is also recognized as the language of the Muslim Minority in Greek Thrace (since 1923 by the Lausanne Treaty) and as such is the main language of education in several minority schools in the area. Although there are no official statistics and the situation is constantly changing, extensive fieldwork allows me to conclude that most Roma from Drosero do not attend these minority schools. With the exception of the families who use Turkish in daily life, Turkish is generally used in the traditional market, the *bazaar*, interactions with the religious authorities, as well as with other Turkish-speaking members of the Muslim Minority (Turks and Pomaks, i.e., traditionally Slavic-speaking populations who are also shifting to Turkish; see Adamou 2010).

Last but not least, Roma children from Drosero receive input in Modern Greek, mainly from outsiders, for example, when they accompany their families to work or to the city, that is, in communicational settings outside of Drosero. Greek, which is the language of the Greek State, is also used in the administration, services outside of the community, and formal education. However, access to schooling is strongly affected by the broader social exclusion and discrimination that Romani communities are confronted to both at the local and at the national level, despite the efforts that are made at the level of the European Union. School attendance among Muslim Roma is low and as a result illiteracy is common among Roma over the age of 40. Though the situation is changing, the youth still frequently drop out before completing middle school.

After this introduction, we now turn to discuss the experimental evidence.

## 2.2. Experimental data on sentence processing of Romani-Turkish

In this section I summarize the results regarding processing costs of mixed Romani-Turkish sentences comprising Turkish non-integrated verbs (Adamou and Shen 2019). The experimental Romani-Turkish data in Adamou and Shen (2019) have been analysed with respect to the literature on codeswitching processing costs, but in this paper I will discuss them in the light of the literature on mixed languages.

Indeed, based on the review presented in the previous sections, one can hypothesize that mixed languages would be processed like unilingual speech, that is, speech with no switching. But how would intermediate stages such as fused lects be processed? Would they be processed similarly to unilingual speech, to codeswitching, or differ from both? Figure 2 illustrates this question.

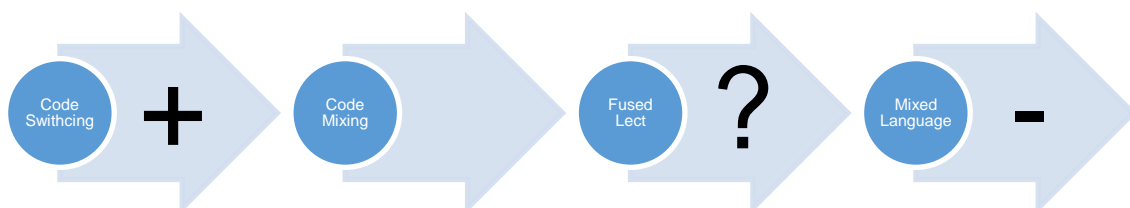




Figure 2. What processing costs for fused lects?

Two research questions have driven the study in Adamou and Shen (2019).

Research question 1: Do mixed Romani-Turkish sentences have higher processing costs than unilingual Turkish sentences? This question can be reformulated as follows: Is conventionalized language mixing (i.e., a fused lect) associated to high processing costs when compared to unilingual speech?

Predictions: In line with several studies on language switching costs, the mixed Romani-Turkish sentences should be associated to higher processing costs than unilingual speech. However, in line with the studies on codeswitching that show no costs and knowing that Adamou and Granqvist (2015) consider Romani-Turkish mixing to be stabilized and predictable there should be no additional processing costs.

Research question 2: Are verbs from Turkish with Turkish morphology processed more slowly when they are in a unilingual Turkish sentence environment or in a mixed Romani-Turkish sentence environment? We can add: Are there any processing costs when this type of morphological non-adaptation is highly conventionalized in the community?

Predictions: Adamou and Shen (2019) predict higher costs in line with the studies that have shown language switching costs. Adamou and Granqvist (2015) hypothesize that the verbs from the contact language in Romani are systematic and therefore predictable. According to this analysis, Turkish verbs should be processed with low costs, similar to unilingual speech. Here, I would like to introduce another prediction based on Myers-Scotton and Jake (2014, 2017) who have argued that non-finite verbs are preferred in codeswitching because they are associated to low costs in production, that is, they require checking at the lexical-conceptual level but not at the level of syntactic and argument structure. Based on this hypothesis, the prediction would be that the morphologically non-integrated verbs that occur in Romani will show high processing costs in production and possibly also in comprehension; it is only the latter that was tested in Adamou and Shen (2019).

To address these two research questions Adamou and Shen (2019) compared the reaction times (RTs), of the Roma participants on two tasks using auditory stimuli: a picture choice (Experiment 1) and a word recognition task in sentence context (Experiment 2). In behavioral studies, RTs correspond to the time participants take to respond to a stimulus. Slower RTs are interpreted as evidence of more difficulties in processing, that is, higher

processing costs. In both experiments, participants were presented with four types of sentences: (a) conventionalized Romani-Turkish mixing involving Turkish verbs with Turkish morphology, (b) all Turkish sentences, (c) sentences with Romani-Turkish codeswitching, that is, with morphologically non-integrated words in lengthy switches (Poplack and Dion 2012), and (d) sentences in Romani with Turkish borrowings, that is, morphologically integrated verbs (Poplack and Dion 2012). Version a offers mixed Romani-Turkish sentences as used in the language of the family and the community. Version b is in the language towards which the shift is currently taking place and that is acquired at home and in the community. Versions d and c are atypical in the community, but they are not ungrammatical since they occur in other Romani communities of the Balkans with which the speakers of Greek Thrace are in contact. Given that sentences of Versions a and b are the most natural in the community, Adamou and Shen (2019) expected that the reaction times would follow the order  $d > c > b \geq a$ . See an example of each version in (4).

Romani in regular font; Turkish in bold; Greek underscored.

(4) ‘The neighbours were having parties very often. They drank, ate (and danced) until late.’

Version a: Mixed Romani-Turkish (as spoken in the community)

a. *e komf-je but seki ker-en-as ylend-ja*

DEF.PL neighbour-PL very often make-3PL-IMPF party-PL

<i>itf-er-di-ler</i>	<i>xa-n-as</i>	<i>but</i>	<b><i>getfi</i></b>	<i>sao</i>	<i>gie</i>
drink-AOR-PST-PL	eat-3PL-IMPF	very	late	every	day

Version b: all Turkish (in the local variety of Turkish)

b. ***komşu-lar ör gün eğlence yap-ıyor-lar-dı***

neighbour-PL every day party make-PROG-3PL-PST

<b><i>İç-er-di-ler</i></b>	<b><i>ye-r-di-ler</i></b>	<b><i>ör</i></b>	<b><i>gün çok</i></b>	<b><i>geç vakt a kadar</i></b>
----------------------------	---------------------------	------------------	-----------------------	--------------------------------

drink-AOR-PST-PL eat-AOR-PST-3PL every day very late time until

Version c: Romani with Turkish codeswitching (not attested in the community)

c. *komfu-lar er gyn ker-en-as eglendze*

neighbour-PL every day make-3PL-IMPF party

*onlar itf-er-di-ler xa-n-as kel-en-as dzi but getfi*

3PL drink-AOR-PST-PL eat-3PL-IMPF dance-3PL-IMPF until very late

Version d: Romani with Turkish borrowings (not attested in the community)

d. *e komf-je er gyn ker-en-as eglendze-a*

DEF.PL neighbour-PL every day make-3PL-IMPF party-PL

*itfki-al-en-as xa-n-as kel-en-as but getfi*

drink-LVM-3PL-IMPF eat-3PL-IMPF dance-3PL-IMPF very late

Sixteen sentences were used, each one constructed in four versions. A norming study for the four versions of the sentences was conducted among Romani speakers from the community. Results confirmed that Versions a and b are the most natural sentences whereas Versions c and d are the most unnatural.

In addition, the stimuli were constructed depending on the language preference of the verbs as observed in the spontaneous conversations. Five sentences included a Turkish verb that is more frequently used in Turkish than in Romani<sup>6</sup> within the community (“Turkish”,

---

<sup>6</sup> Note that speakers of Romani-Turkish generally know the Romani verb as they are in contact with speakers from a variety of Romani dialects, including those that do not borrow extensively to Turkish. However, use of the Romani verb when speaking Romani-Turkish in cases where the Turkish verb is more frequent will be considered as a marked choice. This is particularly straightforward for verbs that have a meaning related to cultural-religious

i.e., ‘to marry’, ‘to return’, ‘to read’, ‘to write’, ‘to think’), six sentences included a Turkish verb that is attested in the corpus once and for which the Romani variant is more frequently encountered (“Romani”, i.e., ‘to come’, ‘to go’, ‘to leave’, ‘to get’, ‘to do/fix’, ‘to put’), and five sentences included a Turkish verb that is used with equal frequency in both Turkish and in Romani (“variable”, i.e., ‘to wait’, ‘to work’, ‘to understand’, ‘to drink’). The motivation for controlling language preference comes from usage-based approaches. Indeed, according to a usage-based approach to language, utterances are composed by smaller units, whether these are words or constructions, which are stored in the speaker’s mind. When the input aligns with a comprehender’s expectations, there are no processing costs, but when it clashes with them, processing costs are entailed (Jaeger and Snider 2013; MacDonald 2013). Predictions were therefore as follows (for the Romani-Turkish version): the verbs preferably used in Turkish will be associated to short RTs (their use is expected and no processing costs will occur), the verbs preferably used in Romani will be associated to longer RTs (their use in Turkish is unexpected and processing costs will occur), and the verbs that are used in either Romani or Turkish will be associated to long RTs (their use in Turkish is plausible but not necessarily expected and processing costs will occur).

### 2.2.1. Experiment 1

Experiment 1 is an online, bimodal picture-sentence matching task with auditory stimuli. It aimed at testing the reaction times of the Romani participants when listening to the various sentences presented above. Thirty seven trilingual Romani-Turkish-Greek speakers participated in the experiment. All were residents of the community of Drosero in Xanthi and had similar low socioeconomic status, and all but one had low education levels, with ages ranging from 13–51 [ $M = 22.59$ ,  $SD = 11.13$ ].

The experiment was conducted on a computer using Open Sesame (Mathôt, Schreij and Theeuwes 2012). Participants had to select the picture that appeared to be more closely related to the meaning of the audio stimuli by pressing a button on the computer.

For the analysis, linear mixed models (lmer) were constructed using the “lme4” package (Bates et al. 2015) in R (R Core Team 2013). The dependent variable was the RT, and the

---

practices, as for example for the verb ‘to marry’, which they will use in Turkish as it is related to a Muslim religious ceremony, while also knowing the verb that other Christian Romani communities are using throughout Greece, which may in turn be a borrowing to Greek.

independent variables were “Language preference” of the verb and “Versions”. The “Subjects”, “Sentences” and “Duration” of audio files were coded as random factors (for more details on the statistical analyses see Adamou and Shen 2019).

Analysis of the results showed that the participants responded the fastest for the unilingual Turkish sentences (version b, see an example in 4b), followed by the Romani-Turkish mixed sentences (version a, see an example in 4a) and the codeswitches (version c, see an example in 4c). Although participants were free to press the button at any moment, RTs indicate that they did not press the button before listening to the Turkish verbs; it was therefore possible to also consider language preference of the verb in the analysis. When taking into consideration the RTs with respect to the language preference of the verbs in the naturalistic corpus, it appeared that the verbs that are more frequently used in Turkish when speaking Romani-Turkish were associated to the fastest responses, whether the sentence was in all-Turkish or in Romani-Turkish. When verbs are more frequently used in Romani but the stimuli included the verb in Turkish, the RTs were slower during processing of the mixed Romani-Turkish sentences than during the processing of the all-Turkish sentences. This indicates that the use of these verbs in Turkish when speaking Romani was unexpected. Finally, the slowest RTs were registered when the Turkish verbs in the stimuli correspond to verbs that are used both in Romani and in Turkish in spontaneous conversations. This indicates that variation in language choice is associated to higher processing costs in comprehension.

### 2.2.2. Experiment 2

This experiment was designed to determine whether the morphologically non-integrated Turkish verbs have higher processing costs when they occur in a Romani environment than in a unilingual Turkish environment. It was an online word monitoring task with auditory sentence stimuli where participants search for a pre-designated target word while listening to language input. This method allows investigating the nature, the position or the context in which the target word is found. Participants were instructed to press the button as soon as they would listen to a specific word (i.e., the target word), in the middle of a sentence. The target word was cut from the original recording and pasted into the instruction sentence. Participants would then listen to a lead-in sentence followed by the critical sentence that included the prime word, which was immediately followed by the target word; see Table 1. Adamou and Shen (2019) considered that if processing a finite verb in two different languages entails costs, then comprehenders would exhibit longer RTs. In contrast, there

would be less or no costs when processing a finite verb in a single language. As can be seen in Table 1, the prime word, a verb from Turkish with Turkish verb morphology, was practically the same in Versions a, b, and c, and differed slightly in Version d. In Versions a, c, d, the target word was the same (Romani verb with Romani verb morphology), but different in the unilingual-Turkish Version b (Turkish verb with Turkish verb morphology, similar to the prime word). A recorded follow-up comprehension question was played in Greek to make sure that the participants processed the sentences and were not merely focused on the search of the target word.

Table 1. Stimulus examples from Experiment 2 (Romani in regular font; Turkish in bold; Greek underscored)

Version	Auditory stimuli for the sentences: 'The neighbours were having parties very often. They drank and danced until late.'	Target word	Prime word
<b>a. Mixed Romani-Turkish</b>	E <b>komfje</b> but <b>seki</b> kenas <u>ylendja</u> . <b>Itferdiler</b> xanas but <b>getfi</b> sao gie.	Xanas	<b>Itferdiler</b>
<b>b. All Turkish (local variety)</b>	<b>Komşular ör gün eğlence yapıyorlardı.</b> <b>İçerdiler yerdiler ör gün çok geç vakt a kadar.</b>	Yerdiler	<b>İçerdiler</b>
<b>c. Romani with Turkish codeswitching</b>	<b>Komfular er gyn</b> kenas <b>eglendze</b> . <b>Onlar itferdiler</b> xanas kenas dži but <b>getfi</b> .	Xanas	<b>Itferdiler</b>
<b>d. Romani with Turkish borrowings</b>	E <b>komfje er gyn</b> kenas <b>eglendzea</b> . <b>Itkialenas</b> xanas kenas but <b>getfi</b> .	Xanas	<b>Itkialenas</b>

Forty nine trilingual Romani-Turkish-Greek speakers participated in this task. Ages range from 13–50 [ $M = 24.10$ ,  $SD = 11.5$ ].

The experiment was again conducted using Open Sesame and statistical analyses were conducted using linear mixed models (lmer) as in Experiment 1. Analysis of the results shows that the RTs in the Romani-Turkish mixed Version a (see example 4a) are significantly shorter than the RTs in the codeswitched Version c (see example 4c) and the RTs in the version with the borrowings (Version d, see example 4d). A similar result is found for the all-Turkish Version b (see example 4b). Crucially, there are no significant differences between the mixed Romani-Turkish Version a and the all-Turkish Version b. This indicates that when preceded by a Turkish verb with Turkish verb morphology the target words are similarly processed whether they are Romani verbs with Romani morphology or Turkish verbs with Turkish morphology. In addition, although the result did not reach statistical significance, participants responded the fastest in mixed Romani-Turkish sentences when the Turkish verb that was the prime word is frequently used in Turkish in natural conversations taking place in Romani. In contrast, when the Turkish verb that was used as a prime word is generally used in

Romani in real life, the all-Turkish sentences were faster than the sentences in mixed Romani-Turkish.

### 2.3. Discussion

Adamou and Shen (2019) set out to answer two research questions, slightly reformulated in the current paper to encompass discussions on mixed language formation and fused lects. The first research question was whether there will be higher processing costs associated to mixed sentences as opposed to unilingual speech. In line with several studies on language switching costs, Adamou and Shen (2019) predicted that the mixed Romani-Turkish sentences should be associated to higher costs than unilingual speech; this is expressed as follows based on the four versions of the experiment:  $d$  (borrowing)  $>$   $c$  (codeswitching)  $>$   $a$  (mixed)  $>$   $b$  (unilingual). Alternatively, in line with Adamou and Granqvist (2015) who consider Romani-Turkish mixing to be stabilized, they predicted similar costs in the mixed sentences as in the unilingual speech;  $d$  (borrowing)  $>$   $c$  (codeswitching)  $>$   $b$  (unilingual)  $\geq$   $a$  (mixed). In Experiment 1 the order of RTs was  $d > c > a > b$ , confirming studies that show processing costs in switching. However, when language preference of the verbs in the natural speech of the community is taken into consideration, it appears that for the verbs that were more frequently Turkish in the corpus, mixed Romani-Turkish was as fast as Turkish unilingual sentences, thus confirming the predictions based on Adamou and Granqvist (2015) for the order  $d > c > b \geq a$ .

The second research question was whether there are costs when processing verb morphology in one language though the rest of the sentence, and neighboring verbs, are in a different language. Again, based on the corpus studies that show conventionalization of the use of Turkish verbs with Turkish verb morphology in a Romani sentence environment, Adamou and Shen (2019) predicted the order  $d > c > b \geq a$ . We could add a second prediction inspired by Myers-Scotton and Jake (2014, 2017) that there should be extra costs when speakers need to check agreement in addition to checking at the lexical-conceptual level. Though Myers-Scotton and Jake (2014, 2017) do not make any predictions in terms of processing costs and though their hypothesis concerns production and not comprehension, it is possible nonetheless to formulate a prediction on this basis. In this case, morphologically non-integrated verbs in Romani should show high processing costs, i.e.,  $d > c > a > b$ . Results of Experiment 2 confirm the first predicted order  $d > c > b \geq a$ , in line with corpus studies



showing that Turkish verbs with Turkish verb morphology are highly conventionalized (Adamou and Granqvist 2015).

Overall, the experimental findings from Romani-Turkish processing lend support to usage-based approaches according to which cognition is shaped by usage (see among others Tomasello 2003 for language acquisition; Construction Grammar in Goldberg 2006; Exemplar Grammar in Bybee 2010; Backus 2015 for codeswitching). Comprehenders therefore anticipate codeswitching based on prior experience (Jaeger and Snider 2013; MacDonald 2013).

Results in Adamou and Shen (2019) confirm that differences in processing boil down to differences in the frequency of use of specific lexical items in natural speech. Although results from Experiment 1 indicate that the comprehension of unilingual speech is less costly than the comprehension of speech involving language switching, a closer look reveals that the sentences that were constructed as closely as possible on the basis of the real-life productions were processed faster when Romani-Turkish mixing occurred than when the sentences were in all-Turkish. This means that cognitive costs depend on the degree to which a switch is expected or unexpected based on previous short-term and long-term language experience. Unusual switches entail longer processing times, most likely due to surprise, and regular switches are processed similarly to sentence stimuli involving a single language. Results from Experiment 2 confirm this analysis as Turkish verbs with Turkish verb morphology are processed similarly whether in a Romani-Turkish mixed sentence or in an all-Turkish sentence. Adamou and Shen (2019) conclude that highly proficient, simultaneous bilinguals do not experience any difficulties in inhibiting one language in order to process a non-integrated word from the other language in a mixed sentence as long as the mixing conforms to established practices in the community.

Let us now turn to discuss these findings with respect to the mixed languages debate. First, I argue that the Romani-Turkish data support the approach according to which fused lects are situated somewhere in between codeswitching (or language mixing) and independent mixed languages (Auer 1999; McConvell and Meakins 2005; O'Shannessy 2012). Indeed, the Romani-Turkish data confirm the existence of a continuum in the conventionalization of language mixing, in this case depending on the degree of conventionalization of specific Turkish verbs, and help refine the category of fused lects as a dynamic stage in the continuum. Specifically, Turkish verbs which are frequently used in Turkish when speaking Romani-Turkish are processed as fast in mixed Romani-Turkish sentences as in unilingual all-Turkish sentences. We could therefore say that when speakers expect some specific verbs to

be in Turkish, such as ‘to marry’ or ‘to think’, based on frequencies in real-life exchanges, then they are processed as though there was no switch, as if the speakers were processing unilingual sentences. This does not mean that they do not know another equivalent, used more frequently among other Romani communities, only that the non-Turkish equivalent is less likely to be used among (Muslim) Roma from Drosero. In contrast, we note higher costs when participants still have the possibility to choose between a Turkish and a Romani verb, such as ‘to work’ or ‘to wait’. This means that the use of those Turkish verbs that still have a Romani equivalent are optional, they are not highly conventionalized yet.

Second, the Romani-Turkish data show that speakers may process the grammatical information of the verbs stemming from two different languages without any additional costs as long as those verbs, together with their morphology, have become conventionalized in the community. The slow RTs for the sentences that were constructed in an unusual manner for the community, for example by using Romani morphology with Turkish verbs, shows that what is important is not so much whether comprehenders will process an integrated or a non-integrated verb but whether they have the habit of listening to integrated or non-integrated verbs. In the terms of usage-based approaches, in language processing it does not make sense to distinguish between grammatical and lexical meaning, as units appear to be stored in memory at various levels of complexity.

In summary, the results of the Romani-Turkish mixing point toward differences in processing depending on frequencies of the language chosen for specific lexical items as established in natural speech. Regarding the continuum, one can therefore reformulate and say that a fused lect is composed of elements that are both highly conventionalized by speakers of a given community and elements that are not yet conventionalized. Figure 3 illustrates this continuum and shows that fused lects are as predicted an intermediate form between codeswitching (or non-conventionalized switches) and mixed languages (or highly conventionalized uses of elements from two sources).

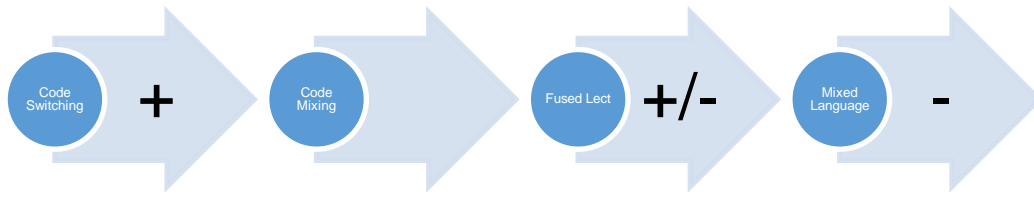


Figure 3. Processing costs for fused lects

### 3. Conclusion

To conclude, in this chapter I have argued that the study of cognitive costs can shed light on the creation of mixed languages. From a more theoretical perspective, I suggested that a fresh look at the mixed language data within usage-based approaches helps understand both the outcome and the process of mixed language creation. In terms of methodology, I hope that the present paper convincingly argued for the merits of the experimental method in the investigation of mixed languages. I consider that very exciting novel perspectives are open to the mixed language debate if we adapt the experimental methods of the research paradigm examining switching costs. It still remains to be seen whether the specific hypothesis formulated in this paper, that mixed languages will have no processing costs, will be borne out by experimental evidence.

More generally, I suggest that in order to better understand the bilingual brain and cognition, we need to take into consideration lesser-documented ways of language mixing involving traditional bilingual communities that might have little exposure to prescriptive and standardized language contexts. The study of these populations raises several methodological hurdles as it may involve participants without any exposure to formal learning for either language, contrasting with most studies on language processing that involve languages with written traditions and exposure to formal teaching. The study conducted by Adamou and Shen (2019) shows that it is possible to overcome those obstacles by working in close collaboration with the communities, after long-term investigations of language practices in the community, and by adjusting our methods for the field.

## Abbreviations

1, 2, 3 first, second, third person; AOR aorist; COM comitative; DEF definite article; FOC focus marker; Q question marker; IMPF imperfective; LVM loan verb marker; NFUT non-future; NOM nominative; OPT optative; PRS present; PRET preterit; PL plural; PST past; SG singular.

## References

- Abutalebi, Jubin & David W. Green. 2016. Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and Cognition* 19. 689–698.
- 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. 2015. A corpus-driven analysis of Romani in contact with Turkish and Greek. In Eivind Torgersen, Stian Hårstad, Brit Mæhlum & Unn Røyneland (eds.), *Proceedings of the 7th International Conference on Language Variation in Europe (ICLaVE 7)*, 1–16. Amsterdam & Philadelphia: Benjamins.
- Adamou, Evangelia. 2016. *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston & Berlin: Mouton de Gruyter.
- Adamou, Evangelia & Amalia Arvaniti. 2014. Greek Thrace Xoraxane Romane (Illustrations of the IPA). *Journal of the International Phonetic Association* 44(2). 223–231.
- Adamou, Evangelia & Kimmo Granqvist. 2015. Unevenly mixed Romani languages. *International Journal of Bilingualism* 19. 525–547.
- Adamou, Evangelia & Xingjia Rachel Shen. 2019. There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism* 23. 53–70.
- Alvarez, Ruben P., Philip J. Holcomb & Jonathan Grainger. 2003. Accessing word meaning in two languages: an event-related brain potential study of beginning bilinguals. *Brain and Language* 87(2). 290–304.
- Auer, P. 1999. From Code-switching via Language Mixing to Fused Lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3(4). 309–332.
- Backus, Ad. 2015. A usage-based approach to codeswitching: The need for reconciling structure and function. In Gerald Stell and Kofi Yakpo (eds.), *Code-switching Between Structural and Sociolinguistic Perspectives*, 19–37. Berlin: Mouton de Gruyter.

- Bakker, Peter. 1994. Michif, the Cree-French mixed language of the Métis buffalo hunters in Canada. In Peter Bakker and Maarten Mous (eds.), *Mixed languages: 15 case studies in language intertwining*, 13–33. Amsterdam: IFOTT.
- Bakker, Peter. 1997. *A language of our own. The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Oxford: Oxford University Press.
- Bakker, Peter & Maarten Mous. 1994. Introduction. In Peter Bakker & Maarten Mous (eds.), *Mixed languages: 15 case studies in language intertwining*, 1–11. Amsterdam: IFOTT.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67. 1–48.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bullock, Barbara, Gualberto Guzman, Jacqueline Serigos, Vivek Sharath & Almeida Jacqueline Toribio. 2018. Predicting the presence of a Matrix Language in code-switching. *Proceedings of the Third Workshop on Computational Approaches to Code-Switching*. 68–75.
- Chan, Mun-Chee, Helen L. H. Chau & Rumjahn Hoosain. 1983. Input/output switch in bilingual code switching. *Journal of Psycholinguistic Research* 12. 407–416.
- Costa, Albert & Mikel Santesteban. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50. 491–511.
- Elšík, Victor & Yaron Matras. 2006. *Markedness and language change*. Berlin & New York: Mouton de Gruyter.
- Friedman, Victor. 2013. Compartmentalized Grammar: The Variable (Non)-Integration of Turkish Verbal Conjugation in Romani Dialects. *Romani Studies* 23(1). 107–120.
- Grainger, Jonathan & Cécile Beauvillain. 1987. Language blocking and lexical access in bilinguals. *Quarterly Journal of Experimental Psychology A*(39). 295–319.
- Green, David W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1. 67–81.
- Green, David W. 2011. Language control in different contexts: the behavioural ecology of bilingual speakers. *Frontiers in Psychology*. <http://journal.frontiersin.org/article/10.3389/fpsyg.2011.00103/full>. Accessed January 15, 2019.
- Green, David W. & Jubin Abutalebi. 2013. Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology* 25. 515–530.

- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gullifer, Jason W., Judith F. Kroll & Paola E. Dussias. 2013. When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology* 4. 1–13.
- Hatzidaki, Anna, Holly P. Branigan & Martin J. Pickering. 2011. Co-activation of syntax in bilingual language production. *Cognitive Psychology* 62. 123–150.
- Henrich, Joseph, Steven J. Heine & Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466. 29–29.
- Ibáñez, Antonio, Pedro Macizo & Maria Teresa Bajo. 2010. Language access and language selection in professional translators. *Acta Psychologica* 135. 257–266.
- Jackson, Georgina M., Rachel Swainson, Ross Cunnington & Stephen R. Jackson. 2001. ERP correlates of executive control during reported language switching. *Bilingualism: Language and Cognition* 4(2). 169–178.
- Jackson, Georgina M., Rachel Swainson, Annie Mullin, Ross Cunnington & Stephen R. Jackson. 2004. ERP correlates of a receptive language-switching task. *The Quarterly Journal of Experimental Psychology* 57A(2). 223–240.
- Jaeger, Florian & Neal Snider. 2013. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127. 57–83.
- Johns, Michael A., Jorge R. Valdés Kroff & Paola E. Dussias. 2019. Mixing things up: How blocking and mixing affect the processing of codemixed sentences. *International Journal of Bilingualism* 23. 584–611.
- Lipski, John M. 2016. Palenquero and Spanish: A first psycholinguistic exploration. *Journal of Pidgin and Creole Languages* 31. 42–81.
- MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology*. <http://dx.doi.org/10.3389/fpsyg.2013.00226>. Accessed January 15, 2019.
- Mathôt, Sebastiaan, Daniel Schreij & Jan Theeuwes. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44. 314–324.
- Matras, Yaron. 2000. A functional-communicative approach. *Bilingualism: Language and Cognition* 3(2). 79–99.

- Matras, Yaron. 2003. Mixed languages: Re-examining the structural prototype. In Yaron Matras and Peter Bakker (eds.), *The mixed language debate*, 151–76. Berlin: Mouton de Gruyter.
- Matras, Yaron, Hazel Gardner, Charlotte Jones & Veronica Schulman. 2007. Angloromani: A different kind of language?. *Anthropological Linguistics* 49(2). 142–184.
- McConvell, Patrick & Felicity Meakins. 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics* 25(1). 9–30.
- 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 Peter Bakker & Yaron Matras (eds.), *Contact languages*, 159–228. Boston & Berlin: Walter de Gruyter.
- Meuter, Renata F. I. & Alan Allport. 1999. Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40. 25–40.
- Mosca, Michela & Harald Clahsen. 2016. Examining language switching in bilinguals: The role of preparation time. *Bilingualism: Language and Cognition* 19. 415–424.
- Moreno, Eva M., Kara D. Federmeier & Marta Kutas. 2002. Switching languages, switching palabras (words): an electrophysiological study of code switching. *Brain and Language* 80(2). 188–207.
- Muysken, Peter. 1997. Media-Lengua. In Sarah G. Thomason (ed.), *Contact languages*, 365–426. Amsterdam: John Benjamins.
- Myers-Scotton, Carol. 1998. A Way to Dusty Death: The Matrix Language Turnover hypothesis. In Lenore Grenoble & Lindsay J. Whaley (eds.), *Endangered languages: Language loss and community response*, 289–316. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 2002. *Contact linguistics, bilingual encounters and grammatical outcomes*. Oxford: Oxford 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.
- Myers-Scotton, Carol & Janice L. Jake. 2014. Nonfinite verbs and negotiating bilingualism in codeswitching. Implications for a language production model. *Bilingualism: Language and cognition* 17(3). 511–525.

- Myers-Scotton, Carol & Janice L. Jake. 2017. Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism* 21(3). 340–366.
- O’Shannessy, Carmel. 2012. The role of codeswitched input to children in the origin of a new mixed language. *Linguistics* 50(2). 305–340.
- O’Shannessy, 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.
- O’Shannessy, Carmel & Felicity Meakins. 2012. Comprehension of competing argument marking systems in two Australian mixed languages. *Bilingualism: Language and Cognition* 15. 378–396.
- Poplack, Shana & Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24(3). 279–315.
- Proverbio, Alice Mado, Giuliana Leoni & Alberto Zani. 2004. Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia* 42(12). 1636–1656.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Accessed October 10, 2013.
- Rusakov, Alexander. 2001. The North Russian Romani dialect: interference and code switching, In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages*, 313–338. Amsterdam & Philadelphia: Benjamins.
- Soares, Carlos & François Grosjean. 1984. Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition* 12: 380–386.
- Tarone, Elaine & Martha Bigelow. 2005. Impact of literacy on oral language processing. Implications for second language acquisition research. *Annual Review of Applied Linguistics* 25. 77–97.
- Thomas, Michael S. C. & Alan Allport. 2000. Language switching costs in bilingual visual word recognition. *Journal of Memory and Language* 43. 44–66.
- Thomason, Sarah G. 1995. Language mixture: ordinary processes, extraordinary results. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: studies in language contact and bilingualism*, 15–33. Washington, DC: Georgetown University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.



Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Wohlgemuth, Jan. 2009. *A typology of verbal borrowings*. Berlin & New York: Mouton de Gruyter.