Modelling contact-induced change in grammar

Jeanette Sakel & Yaron Matras, University of Manchester
(jeanette.sakel@manchester.ac.uk, yaron.matras@manchester.ac.uk)

Abstract
The trivial fact that sociolinguistic circumstances have an impact on language change has been stated over and over again in the context of contact linguistics. Missing from the discussion are, to a large extent, attempts at explicit linking of sociolinguistic constraints on communication in multilingual setting, structural properties of contact-induced change, and the functional value of categories that are prone to change. Attempts at formulating borrowing hierarchies remain constrained by the limited nature of most samples, which normally fail to take into account these three factors systematically while at the same time covering diverse case studies. The present contribution outlines the bottom-up approach to the comparative study of contact-induced change, introducing a database-oriented methodology to describe universal tendencies in the structural outcomes of language contact.

0. Introduction
The outcomes of contact induced language change have been dealt with from many perspectives, such as 1. intensity and impact of contact (code switching, language change, etc.), 2. types of contact situations (linguistic areas, one to one borrowing situations, substrate influence, bilingualism, second language acquisition, etc.), 3. types of contact phenomena (lexical borrowings, grammatical borrowings, calques, borrowing of morphological material), and others such as 4. borrowing hierarchies (which elements are more likely to be borrowed than others). Usually, authors focus on one or only a few of these. Our aim in the context of the research project on Language Convergence and Linguistic Areas (supported by a grant from the Arts and Humanities Research Council) is to develop a systematic approach to analyse and compare the results of contact-induced language change. We wish to take into account the factors involved in different outcomes and to test different types of contact situations, such as linguistic areas. In particular, we wish to focus on the categories affected by contact and compare them to factors such as the social context. In the present paper we discuss our methods of describing contact-induced language change. We will draw on two case studies of languages in contact with Romance languages: the indigenous language Mosetén in
contact with Spanish in the foothills of the Bolivian Andes, and the Kelderash of Romani spoken in Romania in contact with Romanian.

Since our aim is to present a systematic overview of contact-induced change, we must take into account different perspectives of investigating contact. Some definitions of the concepts associated with language contact have been extended or challenged in recent years. These include 1. borrowing vs. code switching, 2. linguistic areas vs. one to one borrowing situations, 3. matter borrowings vs. pattern replication, and 4. borrowing hierarchies. We will briefly discuss each of these.

1. Code switching and long-term borrowing have traditionally been conceived of as unrelated concepts (cf. the overview on code switching and language change given in Backus 2005). In this view, code switching is seen as a mechanism in which two languages are used within the same clause and in the same environment. This is a typical trait of bilingual communication. Borrowing, and with it long-term language change, on the other hand, is seen as the integration of elements from a source language into a recipient language (we use Weinreich’s 1953 terminology). These elements are integrated into the recipient language system and this process usually takes time. In this way, the original definitions of borrowing and code switching differed both in the outcome of the contact, i.e. one versus two languages used in one clause, as well as in the length of contact and “integration”. Most authors draw a clear line between the two when studying contact situations that involve both long-term (propagated) structures and bilingual communication; for example Poplack (1980). The reasons for this can be found in earlier studies of language contact, which often make a sharp contrast between on-the-spot contact phenomena (‘speech’) and integration of features (‘language’). For example Rozencvejg (1976) distinguishes between interference (speech) and convergence (language), treating each of them separately in the two major chapters of his book. His view follows Weinreich’s (1953[1966: 11]) influential work, in which he makes the distinction between interference in speech and interference in language:

“In speech, interference is like sand carried by a stream, in language, it is the sedimented sand deposited on the bottom of a lake. The two phases of interference should be distinguished. In speech, it occurs anew in the utterances of the bilingual speaker as a result of his personal knowledge of the other tongue. In language, we find interference phenomena which, having frequently occurred in the speech of bilinguals, have become habitualized and established. Their use is no longer dependent on bilingualism.”
Weinreich’s distinction is less sharp than Rozencvejg’s (1976) and other interpretations of it. This is well-captured by his analogy of sand, for which we can imagine an intermediate state between being carried by a stream and slowly sedimenting on the bottom of a lake. In recent years, many authors have tried to avoid making a clear distinction between the two (cf. work by Myers-Scotton 1993 2002; Field 2005, etc.); some argue that both concepts can be settled along a continuum of more-or-less integration (cf. Heath 1984, Gardner-Chloros 1995).

2. Other well-established distinctions, of which some have been questioned recently, involve the types of contact situations, such as linguistic areas, one-to one borrowing situations or instances of substrate influence. For example, authors used to draw a clear line between linguistic areas and other contact situations (cf. Thomason 2001, who treats linguistic areas in a separate chapter of her book). In many cases, though, it is difficult to distinguish between areas and borrowing situations. This becomes clear in e.g. Aikhenvald’s (2002) study of borrowing from Tucano (Tucanoan) into Tariana (Arawak) within the linguistic area of the Vaupés, even though many areal traits in Tucano can be attributed to Tariana, or at least Arawak, origins. For this and other reasons, the clear distinction between linguistic areas and one to one borrowing situations has been questioned. Campbell (in press) argues that areas are not fundamentally different from a mere density of shared isoglosses that go back to the diffusion of features between pairs of languages. Similarly, Bakker (in press) shows that within the so-called area of Sri Lanka similar features diffused within pairs of languages, leading to an overall similar profile of some language that have never been in direct contact with each other (cf. also Matras & Sakel in press).

3. One of the distinctions between types of contact phenomena is between borrowed matter (MAT) or calques, i.e. patterns (PAT). This distinction has to do with way elements are affected by the contact (cf. Matras & Sakel in press, Sakel & Matras 2006). ¹ Weinreich (1953[1966: 7]) distinguishes different types of interference: transferred elements vs. interference without outright transfer, which can also be seen in his terminology ‘source’ and ‘recipient’ language for the latter, i.e. MAT, and ‘model’ and ‘replica’ language for the former type of interference, i.e. PAT. Similar approaches are Haugen’s (1950) ‘substitution’ in loanshifts (calques) vs. ‘importation’ (outright borrowing), Gołab’s (1956, 1959) ‘substance’ versus ‘form’, and Johanson’s (1992)

¹ MAT loans usually involve both the morphological material of an element and the structure in which it appears (PAT). Only in few cases do MAT loans only involve the actual matter, cf. Stolz & Stolz (1996, 1997), Sakel (to appear) on function changes in analogy with native markers.
‘global copying’ and ‘partial copying’. Altogether, these concepts have been addressed by many authors in the past, but often they did not play a major role in classifying language contact situations (an exception to this is Heath 1984). Furthermore, a number of studies do not take this distinction into account at all, or only attribute minor importance to it. In this way, Thomason and Kaufman (1988) do not generally distinguish between MAT and PAT loans.

4. Other concepts that have been dealt with within language contact studies include the study of the categories affected and possible borrowing hierarchies. Most of the discussions on this are based on Thomason and Kaufman’s (1988: 74-6) borrowing scale. Their only factor for the degree and type of borrowing is the intensity of the contact; that is more intensive contact would lead to more categories affected in this model. While the intensity of the contact indeed plays an important role in what is borrowed, other studies have shown that it is not the only factor involved (cf. Matras & Sakel in press, Sakel & Matras 2006). Furthermore, Thomason & Kaufman (1988) do not take into account the difference between MAT and PAT borrowings, which appear to have very different distributions in different contact situations. Other discussions of borrowing hierarchies and related concepts have been presented by Moravcsik (1978), Ross (1988) and Lass (1997). In the same way as Thomason & Kaufman’s (1988) borrowing scale, these discussions are top-down in theorising from case studies in the literature. Only few approaches compare sets of first-hand data, such as Curnow’s (2001) examination of the categories affected by contact in the contributions to Aikhenvald & Dixon’s (2001) volume on language contact. Still, dealing with restricted data of very different contact situations, he concludes that small hierarchies may indeed exist, but that it in general is probably impossible to find the constraints on – and factors behind – borrowing. Other studies have shown that small-scale hierarchies indeed can be established, at least for certain categories. In this way, Matras (1998, 2000) shows that there are category-internal hierarchies in borrowing (e.g. but > or > and), and a series of hierarchies are illustrated for a sample of Romani dialects by Elšík & Matras (2006: 343ff.).

It is clear, then, that various factors have to be taken into account when analysing language contact. Classification within strict categories such as ‘area’ or ‘code switching’ is often not possible, nor is it useful in analysing different types of contact situations. A systematic approach, studying first-hand data, could furthermore lead to new generalisations and test existing ones on the mechanisms behind linguistic contact phenomena.
1. Theoretical background, definitions

In our systematic study of language contact, we will take into account various different factors. Most of these deal with the outcomes of language contact. To get a more complete picture of why and how these outcomes differ, we have to begin with studying the initial stages of interference through contact.

At the onset of language contact, subconscious factors often play a role. Matras (1998, 2000, in press) has shown that they relate to speakers’ attempt to reduce the load of processing linguistic operations in discourse, eliminating the need to select among two distinct components of a bilingual individual’s structural repertoire. Occasional, on-the-spot switches can at a later stage become fully integrated into the recipient language; that is undergo propagation in Croft’s (2000) terminology. Social factors play an important role in the process of propagation. Thus, in situations where a highly dominant colonizer’s language is in contact with an indigenous language we usually find propagation of elements from the colonizer’s language into the indigenous language (cf. also Stolz & Stolz 1996, 1997). In bilingual situations with few speakers, on the other hand, propagation is less likely (cf. examples in Matras & Sakel in press). The various types of situations and outcomes of language contact can be settled at various stages within the process, between the very onset, such as on-the-spot switches in a learner’s varieties of a language, and propagation, i.e. established loans in a stable contact situation. In this way, we will not adopt a sharp distinction between code switching and borrowing but view the two of them on a continuum (Gardner-Chloros 1995). In the same way, we do not presuppose a clear-cut boundary between linguistic areas and other contact situations, but will analyse both in the same way and test if such a distinction can be justified on the basis of the categories affected. We will match the categories borrowed up to the social context to see if there are any correlations. Furthermore, we will check for all contact phenomena if they are MAT or PAT loans or intermediate stages of these. This means that we take a bottom-up approach, i.e. look at the actual language data first and from these draw our conclusions about possible borrowing hierarchies.

---

2 Such intermediate stages include the function change of a native element due to phonological similarity with an element in the contact language. In this case, it is difficult to determine whether the loan in question is MAT or PAT.
2. **The procedure**

In this section we outline the method adopted in the project on Language Convergence and Linguistic Areas. We have developed a database to compare data on contact-induced, structural language change in different contact situations. Here we include the different grammatical categories and sub-categories that can be affected by contact, and distinguish various types of social circumstances. The database is designed in the form of a questionnaire for a grammatical description, beginning with general information about the contact situation, followed by information about contact influence in the phonology, typological profile, morphology and different parts of speech, syntactic issues as well as the lexicon. The database consists of ten chapters:

1. Info
2. General
3. Phonology
4. Typology
5. Nominal structures
6. Verbal structures
7. Other parts of speech
8. Constituent order
9. Syntax
10. Lexicon

We include the different concepts discussed above. In this way, we do not distinguish between areas and other borrowing situations in the database. Rather, we provide the possibility to classify a situation as belonging to an area, a borrowing situation, or both, depending on how it is classified in the literature or what the researcher focusses on.

---

3 The database is set up in File Maker Pro.
4 Collaboration with Tadmor & Haspelmath’s ‘loanword typology’ project, looking at the lexical aspects of language contact.
5 We present each contact situation in a separate questionnaire, even though it involves the same language. I.e. Mosetén is both part of a contact situation with Spanish, as well as possibly being part of a proposed Amazonian area (Payne 1996) or Mamoré-Guaporé area (Crevels et al. 2006), having been in contact with other Foothill and Amazonian languages for a long time. Both of these situations are coded in separate questionnaires for Mosetén.
We can then later compare this classification with the actual results from the rest of the database and see if there is a match, and in this way test Campbell’s (in press) hypothesis that areas are mere clusters of one-to-one borrowing situations. We can furthermore compare the different outcomes of borrowing situations to the social factors involved and compare the influence of, for example, official status and language use with the categories that are borrowed to see if there are any correlations. Figure 1 shows how we code the sociolinguistic background and figure 2 shows the encoding of contact languages, both for Mosetén.

Figure 1: Encoding of the sociolinguistic situation (Mosetén):

2.4. Sociolinguistics

Enter any relevant sociolinguistic data, such as language situation, endangerment, age-groups, etc.

Give as much information as you can access.

explanation:

learnt also as L2: when outsiders (i.e. not people living in the village, spouses, linguists)

age groups:

Specify when the language was first (widely) used as a written language

Mosetén is mainly used among friends and in the family. Spanish is highly dominant. Written language materials (in Mosetén) were first introduced in 2000.
### 2.6. Contact languages

Which languages are / have been in contact with this language. 
Please list here all of the contact languages! Include also other dialects of the same language (that are in contact with the variant in question). 
In case of e.g. Spanish as a contact language in South America: list both under current, recent, and older contact - if applicable.

**Current contact:** almost everybody knows the contact language

**Recent contact:** only old(er) people know the contact language, younger people do not speak it any more

**Older contact:** the contact language is known to have existed (by the researcher or the people), but nobody in the community speaks it any more

<table>
<thead>
<tr>
<th>Language(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish, Aymara, Leko, Yurakare, Takana, Trinitario (and other Mojo Arawakan languages), Puquina, Chimane</td>
</tr>
<tr>
<td>Spanish (to lesser degree: Chimane - mutually intelligible with Mosetén)</td>
</tr>
<tr>
<td>Spanish Trinitario, Yurakare, Takana</td>
</tr>
<tr>
<td>Spanish Aymara, Quechua, Leko, Puquina</td>
</tr>
</tbody>
</table>

In the same way as coding for different social circumstances, we also code for different types of interference. In this way, we ask whether an element is taken over as MAT (matter) or PAT (pattern) from the contact language. This distinction is irrelevant in some cases, such as word order, which will always just involve the pattern. Within the expression of other concepts, such as coordination, however, MAT and PAT are two possible outcomes of borrowing. Within these categories (mostly within the chapters on nominal structures, verbal structures and other parts of speech, as well as in some part of the chapter on syntax), we provide the possibility to mark for both MAT and PAT in the database, cf. figure 3. This figure also shows that we can deal with structures for which we do not know if they have arisen through contact influence by marking them as ‘unclear’.
Is is far less straightforward to find out how to code the degree of integration of an element, as well as where on a continuum between borrowing and code switching an element is settled. Adding to this comes the nature of the data, namely that “code switching”, i.e. on-the-spot loans, are much more difficult to quantify than established loans. Rather, most studies investigating these are of a qualitative nature. We also have to consider the different stages of contact, outlined above. Different processes appear at the onset compared to the propagation period, and these would have to be captured as well.

Our solution to this problem is a compromise, opting to be as inclusive as possible. We will include all loans that are more or less ‘integrated’ into the recipient language, i.e. that appear with a certain regularity and that are used by different speakers. In addition, we can mark instances of on-the-spot loans as well. These can be presented as ‘unclear’ cases in the database, followed by an explanation as to why they have been classified this way. This might not always be systematic, and some contributors to the database may classify something as ‘unclear’ which others would include in their list of integrated elements. However, this is a risk worth taking, and the examples and explanations will always provide more information about the structure in question, if this is needed.
3. Case studies: the contact situations of Mosetén/Spanish and Romani/Romanian

The following two case studies on Mosetén/Spanish and Rumanian/Romani are examples of results from contact situations coded in the database.

3.1. The contact situation of Mosetén / Spanish

Mosetén belongs to the small, unaffiliated language family Mosetenan, which is spoken in the foothills of the Bolivian Andes and the adjoining lowland areas of the Amazon basin. The current overview is based on Sakel’s fieldwork on Mosetén of Covendo, one of two major dialects of Mosetén spoken by approximately 800 people (cf. also Sakel 2004).

Mosetén borrows heavily from Spanish. The contact situation is as follows: The Mosetenes have been in contact with speakers of Spanish for hundreds of years. Contact began when missionaries entered the area in the 17th century and there has been substantial contact after permanent missions were established in the 19th century. During those years, bilingualism was common, though many people were still monolingual in Mosetén. Another wave of Spanish influence was the settlement of migrants, mostly from the Andean highlands, introduced as a strategy by the Bolivian government in the sixties and still an ongoing process. Today, the Mosetenes live in close neighbourhood to speakers of Spanish, who often claim a higher status in society and look down on what they consider to be ‘primitive natives’. This situation has led to an immense change in language use – often accompanied by the denial of being of Mosetén origin, when speaking to foreigners. Today, all Mosetenes speak Spanish. There are no monolinguals in Mosetén, rather, many children exclusively learn Spanish. Spanish is used to communicate with outsiders, and at all official purposes, such as school, church, hospital, and for trade. Mosetén is only used orally - mostly among family members and friends, while all writing is done in Spanish. Only very recently, Mosetén has been established as a written language (Sakel 1999, 2001, 2002; cf. also Sakel 2004:50-52), but is not generally used, apart from the attempts of few individuals. Mosetén is and has been – also in a number of other contact situations. For example, it is considered to be part of a Mamoré-Guaporé region (Crevels et al. 2006), and it is located in the

---

6 An exception the work by my main consultant, Juan Huasna Bozo, who is teaching Mosetén at the newly established ‘Universidad de Monseñor Jorge Manrique en Palos Blancos’ in the Mosetén region.
transition zone between the Andean highlands and the (western) Amazonian region (the latter has been proposed by Payne 1990). The analysis of the structures resulting from this contact has only just started – and many aspects of it are still unclear. The Mosetenes are today also in contact with other indigenous languages of Bolivia. They live among speakers of Aymara, Quechua, Trinitario, Yurakare and Tacana, but apart from a few individual speaker’s bilingualism, there is no considerable contact influence from any of these languages. This can be attributed to Spanish as the lingua franca in all types of communication with outsiders. Also the highland settlers, which often belong to Aymara or Quechua speaking communities speak only Spanish with outsiders.

In this way, Spanish is the main contact language for the Mosetenes and the source of many loans in the language. There is both substantial on-the-spot code switching among speakers of Mosetén, as well as integration of a number of elements into Mosetén. For a more detailed discussion of the contact situation and the affected elements cf. Sakel (to appear).

Spanish elements integrated into the Mosetén system include coordinating and subordinating conjunctions, discourse markers, temporal and spatial prepositions and markers expressing restrictions (or delimitation markers). All of these elements can function above the level of the proposition, and have gesture-like functions (cf. Matras 1998): They serve to steer the hearer's processing of propositional content, quite often away from the anticipated course of processing, thus conveying an element of contrast or restriction. Such is the case – generally speaking, and not necessarily in specific connection with Mosetén – with adversative and concessive conjunctions, with phasal adverbs such as ‘already’ and ‘no longer’, with focus particles such as ‘only’, indefinites such as ‘never’, and prepositions such as ‘except’: they all process hearer expectations in respect of a state of affairs, exempting a propositional unit from the set of anticipated units. We shall return to evaluate this tendency below.

With reference to the database questionnaire, most contact phenomena in Mosetén are found in the domains of Other Parts of Speech and Syntax, while there are none or only few phenomena in the other chapters. Going through the categories one-by-one, we’ll give chapter numbers in brackets. Chapters (1) and (2) deal with metadata and with general information on the sociolinguistic situation and have been discussed above already.

There does not seem to be any Spanish influence on (3.) the phonology of Mosetén, apart from loanwords that have not yet been integrated into the phonology of Mosetén. No Spanish sounds are used in native Mosetén words. In the same way, (4.)
the overall typological profile, including categories such as head or dependent marking and alignment, does not seem to have changed through the influence from Spanish. Two constructions in (5.) nominal structures could have arisen through contact with Spanish, both of which involve calquing of the structure (i.e. PAT). The default gender in the system of gender agreement seems to have changed due to Spanish influence. While the default gender for mixed groups used to be feminine, speakers increasingly use the masculine gender in these cases. Along the same lines, plural marking seems to have been adjusted to Spanish patterns. In Mosetén, inanimate objects are usually not marked for plurality but they are increasingly marked as such, as in Spanish. (On the use of Spanish prepositions in Mosetén see below). Contact influence in (6.) verbal structures is limited to the integration of Spanish verbs into the system by native Mosetén verbal markers. There are two markers, only one of which is productive. The productive marker -yi- ‘do’ is used in a similar way in Mosetén as the verbal marker in complex predicates. The unproductive marker -i- is added only to few verbs: viaje-i- ‘to travel’, dewe-i- ‘to owe’, reso-i- ‘to pray’ and fieshta-i- ‘to party’.

Most of the Spanish loans in Mosetén are MAT particle-like elements in the chapter on (7.) other parts of speech. Most of these elements borrowed belong to the category particles and discourse markers. In this way, the Spanish coordinating conjunctions i (from Spanish y, adjusted only in the orthography) ‘addition’, o ‘disjunction’ and pero ‘contrast’ can be used to connect clauses and parts of discourse. The disjunction marker o ‘or’ can also combine phrases. All of the conjunctions appear in the same places as in Spanish, i.e. they are borrowed together with their functions and distributions. They replace native marking of disjunction and contrast by clitics at the end of the second clause, as well as juxtaposition of the clauses to express addition. Sometimes the Spanish and the native elements appear together in double constructions (1):

(1) Tyiñe-tyi’ pero-ki pen’-ki jai’ba-i.⁷
    semi.red-L.M but.E-CO side-CO white-VI.M.S
    ‘It (the peanut) is semi-red, but one side is white.’ STE:VC

⁷ The clitic –ki appears on the first element of the second clause to marker contrastive coordination. Since the conjunction pero is borrowed and added between the clauses, the speaker struggles to identify the first element of the second clause, which can be seen in the marker –ki appearing twice.
Some minor shifts in the usage of the Spanish markers can be noted in Mosetén as compared to Bolivian Spanish. In this way, o is mostly used in constructions expressing alternatives (as o – o, similar to a minor use pattern\(^8\) in Spanish), while mere disjunction is a minor use pattern (in the sense of Heine & Kuteva 2005) in Mosetén. In the same way, the function of pero seems to be somewhat extended in that it not only expresses contrast, but can also mark a change in topic.\(^9\) This extension in function seems to have been motivated by analogy to the native marker -ki, which like pero can be used to express contrast, but which also marks a change in topic in Mosetén.

In the same way, many subordinating conjunctions are borrowed from Spanish, including the complementizer ki (the vowel is adapted phonologically from Spanish que) and several adverbializers. The latter are the si ‘if’, pajki ‘so that’ and the temporal markers hasta, desde and cuando. Also in these cases, the functions of the markers in Mosetén are slightly different from those in Bolivian Spanish. The marker si is only used in conditional clauses expressing alternatives (2), and thus occupies a special position within conditional clauses in Mosetén.

(2) Me’tyi-tyi’ yäe yi-n “si mi rai’s-e’” jäe’mä
DM 1SG say-1SG.O if.E 2SG want-VI-3F.O DM
ji’-chhæ-yi-ti khæei’-si’ phe-ya-k-dye’ o rai’s-e’
CA-know-VI-RE.M.SRF.S-L.F speak-NO or.E want-VI-3F.O
chhi-ban-mi jäe’mä pïñ-i-dye’-in jedye’-jedye’ mö’-yä’
know-again-2SG DM cure-VI-NO-P thing-RD 3F-AD
wiya’-in kихhï tsä’-ïn Köwë’dö’-wë-ïn.
old.man-P go.on.M.S alive-P Covendo-DR-P

‘Thus she said to me “if you want to study our language, or if you want to know about the (native) medicines, there are the old people in Covendo, they are still alive.” STE:JH

A problematic case is the marker pajki, which seems to be a phonologically integrated form of the Spanish purpose marker para que ‘so that’. It expresses both reason and purpose in Mosetén – and thus has an extended function compared to Spanish. If this

---

\(^8\) We adopt this term from Heine and Kuteva (2005).

\(^9\) This might also be a feature inherent to Spanish, cf. Sakel (to appear).
marker indeed is a loan, this extended function could be explained by analogy to native markers that likewise express reason and purpose.

In the same way, the functions of the temporal adverbial clause markers *ashta* (the Standard Spanish orthography is *hasta*) and *kwando* (*cuando* in Standard Spanish) ‘when’ are also extended when appearing in Mosetén. *Ashta* is much more common than *kwando*. Both can denote an endpoint of an event (i.e. ‘until’) (3), as well as the succession of events (i.e. ‘when, after’) (4):

(3) Mi’-khan mi’ bae’-i me’-ki
3M-IN 3M.SGlive-VI.M.S so-CO
keo’-te-in ashta tyaj-ke-te in.
search-VY.3M.O-P until.E find-VK-3M.O-P

‘They searched (for him) where he lived, until they found him.’
WTE:CT

(4) Tyashi si-ti Kallawalla-khan-tyi’-in
first enter.M.S Callawalla-AD-L.M-P
ashta wën-jö-i resya-ya’-in.
until.E move-DJ-M.S church-AD-P

‘The Callawallas are the first to enter when they reach the church.’
WTE:CT

In (4) the meaning is ‘when they come to the church’ and not ‘until they come to the church’, i.e. focusing on the succession of events. *Desde* expresses ‘from, since’ does not seem to be used with an extended function.

Many Spanish discourse markers, that is markers of hesitation, tag questions and the like, are used frequently in Mosetén. These are the tag question *nowe*, from Bolivian Spanish *no ves* ‘don’t you see’ (5) and the hesitation marker *awer*, from Spanish *aver* ‘let’s see’ (6), and *osea* ‘that means, so’. Other, less common markers are *porlomenos* ‘at least’, *siquiera* ‘at least’, *pues* ‘thus, then, well’, *claro* ‘sure’, *claro pues* ‘well, sure’, *bueno* ‘well, sure’, *pues* ‘and then’, and *eso es* ‘that is it!’.

(5) Mö’-nä khä Hernan tipi-ti’
3F.SG-CO well Hernan measure-VT-F.S
mäei’-ya’ jäe’mä Marcelina Duran-tom, nowe?
first-AD DM Marcelina Duran-COM right.E
'And this Hernan, they measured the first time, together with Marcelina Duran, right?’ COE:EC

(6) Alberto, äjj, awer-nä khä, mi’
Alberto EM let’s see E-FO well 3M.SG
jady-i-ti, mi’-nä khä jäe’mä
go.and.come.back VI DT.M.S 3M.SG-FO well DM
chhome’ ītsā-dye-i.
also play.game-NO VI M.S
‘Alberto, well, let’s see, he came [here, performed the action, and went away again], and he was also - uhh - playing games.’ COE:SM

Several Spanish markers of place and temporal deixis are frequently employed in Mosetén as well. These are the prepositions ashta (phonologically adapted from Spanish hasta) ‘until’, desde ‘from, since’, and rarely a ‘to’, en ‘in’, and furthermore the markers nunca ‘never’, siempre ‘always’, kadawes (from Spanish cada vez) ‘every time’ and ai weses (from Spanish a veces). Some of these are parallel to adverbial clause markers.

There are furthermore several Spanish markers that express the opposite meaning or delimitation of some kind. These include embesde (from Spanish en vez de) ‘instead’, ni ‘not even’ and sin ‘without’.

Another construction that could have arisen through PAT interference is the outline of the numeral system, which today is decimal, as in Spanish, but traces in the language point at it originally being a quinary system (cf. Schuller 1917; Sakel 2004:168).

It is much more difficult to analyse the degree to which the (8.) constituent order of Mosetén has been influenced by Spanish, due to the lack of old documents, which could show what the language was like in pre-colonial times or at least at an early stage of contact. Still, a number of constructions seem to have arisen due to Spanish influence, such as the use of prepositions, both in the form of MAT loans in their constructions, as well as through remodelling of native material to fulfil the purpose. Spanish has also influenced other parts of the (9.) syntax of Mosetén, in particular coordination and subordination strategies, that were borrowed together with their markers that are discussed above: MAT-loans usually appear within their Spanish grammatical construction, i.e. coordinating conjunctions appear between the clauses they combine.
Finally, Mosetén has a vast amount to borrowed Spanish (10.) lexicon, in all (lexical) parts of speech.

Summing up, the contact situation is that of a one-to-one borrowing scenario in which the two languages have very different status and prestige. Most structural loans are found at the level of discourse organisation, such as coordinating conjunctions, subordinating conjunctions, deictic markers (of time and space), delimitation markers, and discourse markers. Spanish also seems to have had influence on the grammatical pattern of Mosetén, though this is more difficult to establish.

3.2. The contact situation of Kelderash Romani / Romanian

Romani is the only Indo-Aryan language spoken in Europe since the Middle Ages. It is assumed that the ancestors of the Rom were so-called commercial nomads – specialising in itinerant trades such as metalwork and other craftsmanship, and entertainment. The name ṛom is cognate with the Indian caste-name ḍom, still widespread in India today. Other derivations of the name are found outside India in the self-appellations of the ḍom ‘Gypsies’ of the Middle East, and of the lom of Armenia and the Caucasus. ‘Proto-Romani’ is assumed to have emerged first in Central India, then to have moved to the Northwest. Consequently, it shows both Central archaisms (e.g. Old Indic consonant clusters, Middle Indic conjugation markers) and Northwestern innovations (such as case markers and the past tense formation) (cf. Matras 2002). ‘Early Romani’ is the name given to the immediate precursor of present-day Romani dialects, and is believed to have been spoken within the Byzantine Empire, in close contact with Greek, from around the 10th or 11th century AD and up until the collapse of Byzantium, in the 14th-15th centuries, at which time immigration of Romani groups to other parts of Europe began, leading ultimately to the formation of numerous isoglosses that separate present-day Romani dialects and dialect groups. One of the rather conservative and coherent divisions within Romani is known as the group of ‘Vlax’ dialects (cf. Boretzky 2003), so named after the self-appellation of some of the groups, which in turn derives from their prolonged stay in the Wallachian regions. Characteristic of Vlax, apart from a diagnostic combination of internal archaisms and innovations, is the heavy Romanian-Romance influence on their structure and vocabulary.

One of the well-known Vlax dialects is Kelderash (also Kalderash). As with many of the Romani groups in the Balkans, their name derives from the group’s original
trade specialisation (Romanian căldără ‘kettle-maker’). In various forms, the Kelderash dialect is one of the most widely documented and studied varieties of Romani (cf. Gjerdman & Ljungberg 1963, Boretzky 1994, Matras 1994, Hancock 1995). It belongs to the Northern Vlax dialects, originating in Transylvania, and spoken by a population which in all likelihood was enslaved until the abolition of Gypsy slavery in the Romanian principalities, in the 1860s. Kelderash families began to emigrate from Romania at this point in time, and their communities are found today in urban centres throughout the world. The data considered here, however, is taken from a speaker of Kelderash from southern-central Romania (Piteşti).

Sociolinguistically, Romani is a minority language in contact. All adult speakers of Romani are bilingual (or multilingual), and Romani serves strictly as a language of the extended family and of occasional communication with Roms who are not part of the clan or the community; until very recently, it was strictly an oral language, and not supported by any institutions. Owing to their traditional occupation patterns as a service economy, Gypsies tend to use the majority language in all work-related transactions. The absence of a written or institutional norm in Romani, and the absence of stable communication with other Romani communities, strengthens the dominance of Romanian as the majority language; nonetheless, Romani too has served as a source of lexical vocabulary, primarily for Romanian slang varieties.

In phonology (item 3. on the database questionnaire), the principal shared development is the adoption into Romani of the central vowel /ə/, replacing in some environment an original Romani /e/: žukol ‘dog’ < žukel (cf. also koldărarja ‘kettle-makers, Kelderash’). Major typological features (4.) are largely compatible among the two languages; in part, owing to their shared Indo-European origin, and in part owing to the massive adaptation of Early Romani to Byzantine Greek, and a subsequent ‘Balkanization’ of the language (cf. Matras 1994). Contact influence in the domain of nominal structures (5.) includes the replication of the Romanian plural marker with Romanian-derived nouns (sekrеt–urja ‘secrets’), and its diffusion to both earlier loans (e.g Greek for–urja ‘towns’) and later ones (Slavic čjas–urja ‘hours’), and the incorporation of the analytic comparative/superlative form Romanian mai in adjectives (mai mišto ‘better, best’). Prepositions of Romanian origin include frequent use of în loc dator ‘instead’. Borrowing in the field of verbal structures (6.) is characterised

10 Data collected by Mihaela Zătreanu for the Manchester Romani Project, 2004. For project details and fieldwork methodology see http://www.llc.manchester.ac.uk/Research/Projects/romani/
primarily through the incorporation of several modal verbs (*daštisar*—‘to be able to’, *tržbu*—‘to be necessary’). Romanian loan-verbs are integrated by attaching to the root a loan verb adaptation marker, usually -is-ar-, consisting of a Greek-derived tense/aspect marker -is- and a Romani valency marker -ar- (see discussion in Matras 2002: 128ff.).

The bulk of matter-replications appear in the domain of Other Parts of Speech (7.): Alongside phasal adverbs such as *ynk* ‘still, yet’, *deža* ‘already’, and *či maj* ‘no longer’ we find focus particles such as *numaj* ‘only’ and some time adverbs, such as *domult* ‘for a long time’, *diminjaca* ‘in the morning’, including names of months and days of the week. Among them are also focus particles and connectors:

(7) **niči vov či daštisar-dj-as te putr-el o udar**
    either he NEG can-LOAN-PAST-3SG COMP open-3SG DEF door
    ‘He couldn’t open the door, either’

(8) **niči vov niči lesk-o phral naj len love destul**
    neither he nor his-M brother is.not them.OBL money enough
    ‘Neither he nor is brother has enough money’

The negator *či* (as in *či symas* ‘I was not’) appears to derive from a generalisation of this focus particle, contaminated by the indefinite negator *či* ‘nothing’. Note in (9) that Kelderash Romani borrows the Romanian form of the factual complementizer, *kǎ* (Romanian *cǎ*), expressed elsewhere in Romani generally as *kaj*, but often through a loan:

(9) **dikh-l-em kǎ nas khārɔ**
    see-PAST-1SG COMP was.not at.home
    ‘I saw (that) he was not at home’

Other grammatical function words of Romanian origin include discourse markers and utterance modifiers:

(10) **bajda te anklj-as avri**
    let’s COMP exit-1P out
    ‘Let’s go out’
(11) **atunči** le la po bango vast kaj dujto intersekacija
    then take it at crooked hand at-DEF second intersection
    ‘Then take a left turn at the second intersection.’

Also belonging to Other Parts of Speech is the temporal indefinite *yntodjauna* ‘always’, and the temporal negative indefinite ‘never’, a hybrid expression containing the Romani negative indefinite *či*, the Romani numeral *ek* ‘one’, and the Romanian temporal unit expression *data* ‘time’:

(12) *či ža-u či ek data* kaj pjaca te tin-au manřo
    NEG go-1SG NEG one time to.DEF market COMP buy-1SG bread
    ‘I never go to the market to buy bread.’

The adoption of word forms for indefinite pronouns also includes *nište* ‘anything’ as well as of the indefinite marker *vare-* (in combination with Romani interrogatives, e.g. *vare-kon* ‘someone’, *vare-so* ‘something’ etc.).

In constituent order (8.) and the syntax of complex constructions (9.), both Romani (generically, that is, Early Romani and the southeastern European dialects of the language) and Romanian share characteristic features of the Balkan linguistic area, such as flexible VS/SV order, finite adverbial and relative clauses introduced by subordinating conjunctions, finite modal complements (instead of infinitives), distinct conjunctions introducing factual and non-factual complements, and more. Finally, a massive amount of lexical material (10.) is borrowed from Romanian into Kelderash Romani, including basic nouns such as *fejastra* ‘window’, *marja* ‘sea’, *vorba* ‘word’, *lumina* ‘light’, *animalo* ‘animal’, basic verbs such as *troil* ‘live’, *skrisarel* ‘write’, *jubisarel* ‘love’, and basic adjectives such as *lungo* ‘long’ and *skurto* ‘short’.

4. Conclusions
Our goal was to discuss how to model a comparative description of contact-related change, in order to arrive at generalisations about mechanisms of language contact. We have presented a method which has the following features:
1) Consideration of the sociolinguistic background, and functions of the language in the community
2) Tagging of types of contact: area, one-to-one
3) Review of all possible structures affected by contact, by word class and category; rendering a kind of grammatical description from the contact perspective
4) Tagging of contact influences for different structural types, MAT and PAT
5) Representation of a continuum of spontaneous and established contact influences
6) Employment a database structure for ease and efficiency of comparison

The database is currently in the final stages of setup, with invitations to specialists in contact situations to contribute data, it will eventually go online and be accessible for researchers, constituting a first and unique systematic, modifiable cross-linguistic sample of contact influences.

We have chosen here to discuss, briefly, two of our sample idioms that are under the influence of Romance languages, Mosetén and Kelderash Romani. Their histories differ considerably: the first is a colonized language, the second a migratory diaspora language. Nonetheless, we find similarities in the domains of use, and in the role of the respective dominant contact language. All speakers are bilingual, the dominant language serves for all contacts outside the community and with institutions, while the group language remains oral, and a symbol of in-group identity. There is therefore motivation to hold on to the community language, and acceptance of bilingualism.

The structural profiles of contact are not dissimilar, either (Table 1). Most prominent are loans in the domain which we termed ‘Other parts of speech’, especially discourse markers and connectors, temporal expressions and focus particles – the inventory characterised by Matras (1998) as ‘utterance modifiers’ triggering ‘monitoring-and-directing’ operations in discourse interaction. A hypothesis lends itself, on the basis of just this one comparison, that oral languages under the influence of dominant, institutionalised majority idioms will tend to adopt first those uninflected function words that play a major role in structuring the utterance at the level of the discourse interaction. Note the prominence, among the borrowed items in both languages, of discourse markers, temporal indefinites, and complementisers. Mosetén goes further in its adoption of conjunctions, both coordinating and subordinating, presumably due to the need to adjust its clause combining strategy to the Spanish model, while Kelderash Romani enters into contact with Romanian already possessing similar clause combining patterns, and relying on native conjunctions. Nonetheless, the
pattern of conjunction borrowing is prevalent in other Romani dialects in contact with other languages (cf. Matras 2002: 201-202), and Kelderash manifests itself in this regard as rather conservative. On the other hand, compared with Mosetén, Kelderash Romani borrows more in the areas of indefinites, phasal adverbs, and focus particles, as well as in the area of modality.

Noteworthy are some other striking similarities between the two cases. In prepositions, both borrow the delimitation form ‘instead’; in nominal structures, the representation of plurality is affected; in verbal structures, valency or ‘verb-ness’ markers are used to adapt loan-verbs; and there are no visible contact phenomena in the domain of morphological typology. Equally striking is the absence of borrowing in certain domains, such as personal pronouns, location deixis, and demonstratives. This brings to our attention the supposed proneness of unbound ‘function words’ to borrowing. In fact, in close examination it appears that it is not the structural class as such that attracts borrowing, but only some of its sub-categories, and so borrowing follows a semantic-functional hierarchy, capturing first and foremostly those grammatical devices that process expectations: anticipated links between propositions, time phases, sets, and qualifications of propositions. In an evaluation of the Romani sample, Elšík & Matras (2006: 385-386), drawing on Matras (1998), argue that borrowing targets firstly those grammatical devices that are employed to process instances of potential tension between the message conveyed by the speaker, and the hearer’s expectations. Take for instance the preposition ‘instead of’, which exempts an item from a presupposed set, or the contrastive marker ‘but’ or focus particle ‘only’, which convey similar exemptions. In such instances of mental tension surrounding the processing of the utterance, there is an advantage to the bilingual speaker to improve the efficiency of retrieval of the correct expression by eliminating the need to select between sub-components of the linguistic repertoire – the two ‘languages’. The result is ‘fusion’ – the non-separation of the two linguistic systems around the device in question. It is this which, in diachronic perspective, we call ‘grammatical borrowing’.

The cross-linguistic study of grammatical (and lexical) borrowing is still in its infancy, despite the vast amount of literature dedicated to the subject. We hope to have outlined a model that will enable to pursue this study in a more systematic and transparent fashion, so that solid cross-linguistic generalisations may be arrived at soon, opening new avenues toward interpreting the role of structure, semantic and pragmatic function, and social constraints in the complex process known as contact-induced change.
Table 1: Grammatical contact phenomena in Mosetén and Kelderash Romani

<table>
<thead>
<tr>
<th>Questionnaire chapter</th>
<th>Sub-category</th>
<th>Spanish in Mosetén</th>
<th>Romanian in Kelderash Romani</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information</td>
<td>Fieldwork (Sakel)</td>
<td>Fieldwork (Manchester Romani Project; Zatreanu)</td>
<td></td>
</tr>
<tr>
<td>2. General</td>
<td>MOSETÉN &lt; MOSETÉNAN (isolate); family &amp; community oral language; only as L1; all speakers bilingual</td>
<td>KELDERASH ROMANI &lt; VLAX ROMANI &lt; ROMANI &lt; INDO-ARYAN; family &amp; community oral language; only as L1; all speakers bilingual</td>
<td></td>
</tr>
<tr>
<td>3. Phonology</td>
<td>Adoption of central vowel /a/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. (Morphological) typology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Nominal structures</td>
<td>Prepositions: delimitation <em>embesde</em> ‘instead’, <em>sin</em> ‘without’</td>
<td><em>in loc də</em> ‘instead’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Prepositions: temporal <em>ashta</em> ‘until’, <em>desde</em> ‘from, since’,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>General: masculine as default; explicit plurality plural marker -urja; adjective comparison maj</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Verbal structures</td>
<td>Verbal structures: loan verb integration with native verbalising markers with native valency markers and Greek-derived tense/aspect markers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Verbal structures: other modal verbs <em>dašisar-</em> ‘can’, <em>trob-</em> ‘must’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Other parts of speech</td>
<td>Discourse markers: sequential <em>pues</em> ‘then’</td>
<td><em>atunči</em> ‘then’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discourse markers: other <em>aver, nowe, bueno</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coordinating conjunctions <em>i</em> ‘and’, <em>o</em> ‘or’, <em>pero</em> ‘but’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complementiser general: <em>ki</em> factual: <em>kə</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adverbial subordinators <em>kwando</em> ‘when’, <em>si</em> ‘if’, <em>pajki</em> ‘so that’, <em>ashta</em> ‘until’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indefinites: <em>nunca</em> ‘never’, *siempre’</td>
<td><em>či ek data</em> ‘never’, <em>yntodjauna</em></td>
<td></td>
</tr>
<tr>
<td>7. Other parts of speech</td>
<td>Indefinites: other</td>
<td>vare/- or-, nište ‘anything’</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>7. Other parts of speech</td>
<td>Phasal adverbs</td>
<td>ynkо ‘still, yet’, či maj ‘no longer’, deža ‘already’</td>
<td></td>
</tr>
<tr>
<td>7. Other parts of speech</td>
<td>Focus particles</td>
<td>nici ..(n)či ‘(n)either ...(n)or’, numaj ‘only’</td>
<td></td>
</tr>
<tr>
<td>8. Constituent order</td>
<td>Position of Spanish prepositions</td>
<td>Compatibility of both systems with Balkan ‘type’ (areal convergence)</td>
<td></td>
</tr>
<tr>
<td>9. Syntax</td>
<td>Partial adoption of Spanish format in clause combining</td>
<td>Compatibility of both systems with Balkan ‘type’ (areal convergence)</td>
<td></td>
</tr>
<tr>
<td>10. Lexicon</td>
<td>Temporal expressions</td>
<td>kadawes ‘every time’</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>damult ‘for a long time’, diminjaca ‘in the morning’, days of the week, months</td>
<td></td>
</tr>
<tr>
<td>10. Lexicon</td>
<td>Lexicon: other</td>
<td>In most semantic domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>In most semantic domains</td>
<td></td>
</tr>
</tbody>
</table>

**References**


Campbell, Lyle. in press. ‘Areal linguistics: a closer scrutiny’ in Yaron Matras, April


Cambridge University Press.


Sakel, Jeanette & Yaron Matras (2006). Prototypical linguistic areas and
borrowing hierarchies, toward an integral investigation of linguistic areas. Paper presented at the workshop on ‘Current issues in areal typology’ at the meeting of the DGfS (Deutsche Gesellschaft für Sprachwissenschaft) in Bielefeld, Germany, 22.2.2006.

Schuller, Rudolph. 1917. Introduction to Benigno Bibolotti’s Moseteno Vocabulary and Treaties from an unpublished manuscript in possession of Northwestern University Library. Evanston and Chicago: Northwestern University.


List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>adessive</td>
</tr>
<tr>
<td>CA</td>
<td>causative</td>
</tr>
<tr>
<td>CO</td>
<td>contrastive marker</td>
</tr>
<tr>
<td>COE</td>
<td>conversation ex.</td>
</tr>
<tr>
<td>COM</td>
<td>comitative</td>
</tr>
<tr>
<td>COMP</td>
<td>complementizer</td>
</tr>
<tr>
<td>CT</td>
<td>Cleto Tahe</td>
</tr>
<tr>
<td>DJ</td>
<td>assoc. motion –kho-</td>
</tr>
<tr>
<td>DEF</td>
<td>definite article</td>
</tr>
<tr>
<td>DM</td>
<td>discourse marker</td>
</tr>
<tr>
<td>DR</td>
<td>downriver relation</td>
</tr>
<tr>
<td>DT</td>
<td>assoc. motion –ti-</td>
</tr>
<tr>
<td>E</td>
<td>Spanish loan</td>
</tr>
<tr>
<td>EC</td>
<td>Esteban Condo</td>
</tr>
<tr>
<td>EM</td>
<td>emphatic marker</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>FO</td>
<td>focus</td>
</tr>
<tr>
<td>IN</td>
<td>inessive</td>
</tr>
<tr>
<td>JH</td>
<td>Juan Huasna</td>
</tr>
<tr>
<td>L</td>
<td>linker</td>
</tr>
<tr>
<td>LOAN</td>
<td>loan verb marker</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>NEG</td>
<td>negation marker</td>
</tr>
<tr>
<td>NO</td>
<td>nominalisation</td>
</tr>
<tr>
<td>O</td>
<td>object</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>P</td>
<td>plural</td>
</tr>
<tr>
<td>PEM</td>
<td>personal pronoun</td>
</tr>
<tr>
<td>PAST</td>
<td>perfective tense</td>
</tr>
<tr>
<td>RD</td>
<td>reduplication</td>
</tr>
<tr>
<td>RE</td>
<td>reflexive, reciprocal</td>
</tr>
<tr>
<td>RF</td>
<td>logophoric reference</td>
</tr>
<tr>
<td>S</td>
<td>subject</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SM</td>
<td>Sacarias Misange</td>
</tr>
<tr>
<td>STE</td>
<td>spoken text example</td>
</tr>
<tr>
<td>VC</td>
<td>Vitoriano Chairique</td>
</tr>
<tr>
<td>VI</td>
<td>verbal stem –i-</td>
</tr>
<tr>
<td>VK</td>
<td>verbal stem –ki-</td>
</tr>
<tr>
<td>VT</td>
<td>verbal stem –ti-</td>
</tr>
<tr>
<td>VY</td>
<td>verbal stem –yi-</td>
</tr>
<tr>
<td>WTE</td>
<td>written text example</td>
</tr>
</tbody>
</table>