4 Mixed languages and language intertwining

Peter Bakker and Pieter Muysken

4.1 Introduction

This chapter is about languages in which the morpho-syntax of one language is matched with the vocabulary of another language. They are not creoles or pidgins in the strict sense, but they may shed light on the genesis of these languages as well. We will begin with a few examples.

Angloromani is the language of tens of thousands of Gypsies in Britain, the United States and Australia. The grammar (phonology, morphology, syntax) is English, but the words are overwhelmingly of Gypsy or Romani (Indic) origin. The following example, with Romani elements italicized, is from Smart & Crofton (1875: 218-221). In this example, and in the following, we will give the examples in the mixed language as well as in the two source languages.

(1) **Palla bish besh-es apopli** the Beng wel'd and **pen’d:**

after 20 year-s again the Devil came and said

*Av* with **man-di.**

come with **me-DAT**

Palla bish besh-aw apopli o Beng vi-as. **Yov pen-das:**

after 20 year-PL again the Devil come-3.PA. **He say-3.PA**

*Av* man-tnsa.

come me-with

‘After twenty years the Devil came back and said: come with me.’

When Angloromani was first discussed by language contact specialists, it was related to creole languages (Hancock 1970), which are also said to have a different origin of lexicon and grammatical system.

Media Lengua (Ecuador) is another example. Virtually all the lexemes (italicized) are of Spanish origin, whereas the grammar is almost identical to that of the local Quechua.

(2) **Yo-ga awa-bi kay-mu-ni.**

I-to water-LO fall-CI-1
Muysken (1981b) claimed that the lexical shapes of Quechua content words have been replaced by Spanish ones.

Ma’a or Mbugu (Tanzania) posed some problems for language classification, as it was sometimes classified as Bantu and sometimes as Cushitic (Goodman 1971). The grammatical system is virtually identical to the neighboring Bantu languages, but the lexicon of content words is overwhelmingly of Cushitic origin (Thomason 1983). In the following example, Bantu elements are italicized (Mous 1994a):

(3) *hé-lo mw-a-giru é-se-we Kimweri, diláó w-a*  
16NC-have INC-CON-big INC-call-PA=PF King Kimweri king INC-CON  
yá idi lá Lusótó.  
this 5NC-land 5NC-CON Lushoto

‘There was an elder called Kimweri, king of this land of Lushoto.’

Thomason (1983) and Thomason & Kaufman (1988) took this language as being the result of massive grammatical replacement: the original Cushitic grammar had been replaced by Bantu grammar.

We have now looked at three cases of languages with a lexicon originating from a language different from that of the grammatical system. We propose one single general process as being responsible for the genesis of these three languages and the others like them. We call this language intertwining, which should be taken to be a type of language genesis different from cases like creolization or pidginization, and also from lexical borrowing and language shift. Language intertwining is a process which creates new languages which have roughly the following characteristics:

(4) An intertwined language has lexical morphemes from one language and grammatical morphemes from another.

Media Lengua is spoken by people whose parents are Indians who spoke only Quechua, therefore the Spanish lexical elements must have been introduced into Quechua. In the case of Angloromani, the speakers are Gypsies who formerly spoke Romani; therefore the English (grammatical) elements seem to be the intruding element. The same holds for the Ma’a
speakers, who are ethnohistorically Cushitic, and here as well the grammatical elements are what appears to have been added. While historically there would seem to be two different processes, they still lead to similar results. We claim that on a more abstract level, they are one and the same: intertwining a grammar and a lexicon from different sources. There is no reason to suppose that two entirely different processes would lead to the same type of mixed language, viz. massive grammatical replacement in the case of AngloRomani and Ma’a (as Thomason & Kaufman argue) and relexification in the case of Media Lengua.

Before we can explore the process of intertwining, we will need to look at several intertwined languages in more detail.

4.2. Media Lengua (Quechua grammar, Spanish lexicon)

The variety of Media Lengua (ML, lit. 'half language' or 'halfway language') described here is spoken natively by up to a thousand people in Central Ecuador (Muysken 1979, 1981b, 1988b, in press a.). Further examples of Media Lengua utterances are given in (5) and (6), with the (b) examples giving the regional Quechua equivalent, and the (c) examples the regional Spanish equivalent. A first example is:

    one favor-AC ask-FN-BN come-PR-I
    (Media Lengua)

    one favor-AC ask-FN-BN come-PR-I
    (Quechua)

c. Veng-o para ped-iR un favor.
    come-1SG to ask-INF a favor
    (Spanish)

'It I come to ask a favor.'

It is clear that (5a) has resulted from combining the phonological shapes of the stems in (5c) with the grammar and lexical entries in (5b). Thus shuk is replaced by unu, mana- by pidi-, etc. Several things should be noted. First, we get an emphatic form of the indefinite article in Media Lengua, unu, rather than Spanish unemphatic un. Second, the Spanish irregular verb form vengo appears in a regularized stem form bini. Third, the Quechua rule voicing the accusative case marker -ta to -da after fabur has not applied in Media Lengua; Quechua dialectological evidence suggests that this is a recent rule. Fourth, the Spanish forms have been adapted phonologically to Quechua; mid vowels have been replaced by high vowels. Quechua word order and morphology have been retained.
Mixed languages and language intertwining

(6) a. Kuyi-buk yirba nuwabi-shka. (Media Lengua)
cavia-bN grass there.is.not-SD
b. Kuyi-buk k'iwa illa-shka. (Quechua)
cavia-bN grass there.is.not-SD
c. No hay hierba para los cuy-es. (Spanish)
NEG there-s grass for the-pl cavia-pl
'There turns out to be no grass for the guinea pigs.'

Note that the Quechua word kuyi 'guinea pig' appears in the local Spanish as well. The Media Lengua verb maintains the Quechua-specific 'sudden discovery tense' marking -shka. The Quechua negative existential verb stem illa- has been replaced by a newly formed 'frozen' stem nuwabi-, derived from Spanish no and haber 'have'. The Spanish verb 'have' has an impersonal form hay which also has existential meaning.

To summarize, Media Lengua is essentially the product of replacing the phonological shapes of Quechua stems with Spanish forms, maintaining the rest of the Quechua structure. The Spanish forms chosen have undergone regularization and adaptation to Quechua morphophonology. Media Lengua is conservative in sometimes reflecting earlier stages in Quechua pronunciation. It is not made up on the spot every time it is spoken. The occurrence of Spanish strong alternants (cf. Spanish uno), frozen composites (cf. nuwabi-), etc. is an indication that we do not have a simple process of vocabulary replacement here.

What is peculiar about Media Lengua is not so much that it contains Spanish words, but that almost all Quechua words, including all core vocabulary, have been replaced. All Quechua dialects borrowed heavily from Spanish, up to roughly 40%, but there are no dialects which borrowed more than 40%. Thus there is an enormous gap between the 40% of hispanicized Quechua dialects and the over 90% of Media Lengua.

Muysken (1981b) proposed that Media Lengua had come into being via a process of relexification: the replacement of the phonological shape of a root of one language (Quechua) by a root with roughly the same meaning from another language (Spanish). There were several reasons for supposing a process of borrowing of Spanish vocabulary into Quechua rather than the borrowing of Quechua affixes and grammar into Spanish. In the first place, borrowing of roots is much more common than borrowing of affixes – the latter is hardly possible without preceding massive lexical borrowing. Second, the Spanish roots sometimes have semantic properties identical to the replaced Quechua root. Thus Media Lengua sinta- 'live, exist, be in a certain place, sit' has the whole range of meanings of Quechua tiya- rather than that of Spanish sentar(se) 'to sit, sit down'.

Media Lengua is not the product of an interlanguage frozen and fixed, resulting from communicative needs; it is not used with outsiders, but it isn't a secret language either.
Rather it is an ordinary, day to day, community-level form of communication.

The Media Lengua-speaking communities studied here are located on the fringe of a Quechua-speaking area, to which the community historically belonged. Due to its geographical situation, and due to the necessity and opportunities for its inhabitants to make frequent trips to the capital to look for work, the community has come to be culturally differentiated from neighboring areas to the extent that its people find it necessary to set themselves apart from the neighbors. Between 1900 and 1920, the capital of Quito expanded enormously, and many Indian construction workers from the relevant area were employed there. It is possibly this group of migrant construction workers started the process of relexification. Now, however, many people who never worked in construction speak Media Lengua as well.

4.3 Michif (Cree grammar, French lexicon)

Michif is a mixture of Cree and French which is spoken by fewer than 1000 people of Cree and French ancestry in Western Canada and in North Dakota and Montana, United States. Cree is the name of an Amerindian nation and their language, which belongs to the Algonquian language family. The Michif have considered themselves to be members of the Métis or Michif (‘Mestizo’) Nation since around 1800. The word Michif is an anglicized spelling of the Métis pronunciation of the French word Métif.

The Michif language differs in a way a from the cases discussed here: it is often described as having Cree verbs and French nouns (cf. Thomason and Kaufman 1988, Rhodes 1977). Bakker (1992) calculates that between 83 and 94% French nouns and between 88 and 99% Cree verbs are used, depending on the community. It has been argued, however, that the real dichotomy between the two language components in Michif is that the grammatical system is Cree and the lexicon French (Bakker 1990, 1992). The verbs are claimed there to consist of only bound morphemes and are therefore always in Cree (cf. Goddard 1990). It is therefore a case of language intertwining, be it deviant because of the polysynthetic structure of Cree. Some examples of Michif from Bakker’s field notes can be found in the sentences:

(7) Ki-nipi-yi-wa son frère aspin kâ-la-petite-fille-iwi-t.
    st-die-OB-3SG his/her brother since comp-the-little-girl-be-3SG
    ‘Her brother died when she was a young girl.’

(8) John ki-wéyisim-éw Irene-a dans sa maison kâ-pihtikwé-yi-t.
    John PA-lure-3S.30 Irene-OB in 3SG house comp-enter-OB-3SG
    ‘John lured Irene into his house.’
In these examples we see that all the verbs are Cree. All the nouns (and nominal modifiers such as possessives) are French, and the proper noun 'Irene' in (8) has a Cree nominal marker for obviation (a third person noun phrase that is not a topic).

In Michif, nouns, numerals, definite and indefinite articles, possessive pronoun, some adverbs and adjectives are from French, but demonstratives, question words, verbs and some adverbs and (verb-like) adjectives are from Cree. Whereas French words can appear with Cree morphology, the opposite is not the case. Both languages retain their own phonological systems.

### 4.4 Ma’a (Bantu grammar, Cushitic lexicon)

Ma’a is spoken by approximately 12,000 people in the Usumbara district in north-eastern Tanzania, close to the Kenya border (Goodman 1971). It has been little studied so far in any detail. Mous (1994, to appear) is doing systematic fieldwork on the language. While there is still some disagreement about the specific provenance of certain elements, the structure of Ma’a is clear in general. Most of the lexicon is Cushitic but the noun class and verbal inflection system is Bantu, as is the word order.

Ma’a has been looked at from the perspective of language contact studies by several researchers, notably Goodman (1971) and Thomason (1983). Goodman hypothesizes that it came about through gradual convergence of the Bantu and Cushitic languages spoken in a bilingual community, and that there may have been simplification as well. He does not exclude imperfect second language learning of both component languages by yet a third group (1971:252). Thomason (1983) claims that a scenario by which there was massive interference from a Bantu language in a Cushitic language is the most plausible one. The circumstances were such that speakers attempted to maintain their Cushitic language while being under intense cultural pressure from a Bantu language. Thomason rejects a scenario by which there was massive interference from a Cushitic language B into a Bantu language A (1983:220-1), since then ‘we would again expect Ma’a basic vocabulary to be primarily of Bantu origin.’ Both analyses are quite different from what we claim.

### 4.5 Kröjo or Javindo (Low Javanese grammar, Dutch lexicon)

Kröjo is the language spoken by the descendants of European (mainly Dutch) fathers and Javanese mothers (formerly?) spoken in Semarang, Java. The discussion in this section is based on De Gruiter's monograph (1990) on the language, of which the author was a speaker in his youth. Kröjo is a derogatory term in Javanese for a person of mixed European – Indonesian ancestry (De Gruiter 1990: 17) and this is also used for this mixed language by
the people who spoke it. De Gruiter prefers his coined term ‘Javindo’. In addition to Javanese-related Javindo, there are other intertwined languages in Indonesia as well, which have Malay as their base.

An example of this language (from De Gruiter 1990) is (9); Dutch elements are italicized.

(9) Als ken-niet, ja di-ken-ken-a, wong so muulek kok sommen-nja.
if can-not, IM PAS-BL=RED-IR, (EM) for (so) difficult those sum-(PL)-DEM
Nak ora isa, ya di-isa-isa-a, na wong angèl temen kok suwal an-è ekkok m-en-nja.
(Javanese)

Als het niet gaat, stel dan alles in het werk opdat je wel kunt, want die sommen zijn zo moeilijk.
(Dutch)
‘If it is not possible, try to get them done, for the sums are hard.’ (-nja is a Malay demonstrative, -ém its Javanese equivalent.)

De Gruiter considers it ‘a language which was structurally Javanese with as many Dutch words as possible.’ (1990: 23).

4.6 Intertwined Romani languages

In Europe there are a number of examples of intertwined languages, mostly spoken by nomadic peoples such as (certain) Gypsies and Irish Travellers. Hancock first suggested a contact origin for Angloromani (1970, 1984a,b) and Shelta (Hancock 1974, 1984c), and Boretzky was the first who compared three languages with Romani (Gypsy) lexicon and grammars from other languages (Boretzky 1985). Bakker & Cortiade (1991) contains several studies on Romani mixed languages (Basque, Greek and Armenian) as well as a general article. Boretzky and Igla (1994) is another overview.

Romani is a language of Indic origin, most closely related to the languages of northwest India such as Punjabi. The ancestors of the Gypsies left India around the year 1000 and they arrived in Europe in the 14th century, where the language split into dialects with varying degrees of influences from Balkan languages, especially Greek, Slavonic and (in some varieties) Rumanian. Some of the Romani dialects, however, have lost the Indic and Balkan grammatical features and now have only the Romani lexicon but a different grammatical system. The cases of these Romani intertwined languages identified thus far involve Greek, English, Danish, Swedish, Norwegian, Occitan, French, Spanish, Brazilian Portuguese, Catalan, Basque, Turkish, and Western Armenian. Not all of these are well documented, however. In this section we will briefly deal with a number of the better documented Romani varieties, discussing in some more detail Basque Romani, Swedish Romani and
Spanish Romani, since these three languages are typologically rather different. Basque is a language with complex morphology, partly agglutinative and partly fusional, both in the noun and the verb. Swedish is a language with very little morphology and Spanish is an intermediate case, having fusional morphology in verbs and nouns.

Below are examples from Swedish Romani (Hancock 1992), Spanish Romani or Caló (Bright 1818), and Basque Romani (Ibarguti 1989).

(10) Vi *trad-ar* to *får on* en vaver *divus*.  
    (Swedish Romani)
    Vi *åk-er* till stan *en annan* dag.  
    (Swedish)
    Dž-as ka o *foros* vaver dives.  
    (Romani)
    *We('ll) go to town another day.*

(11) Se ha *endiñ-ado* el *parné* a la *chai*.  
    (Spanish Romani)
    Se ha da-do el dinero a la muchacha.  
    (Spanish)
    E cha-ke dend-ile o love.  
    (Romani)
    *The girl was given the money.* (parné is r. for ‘white (things)’, here a metaphor for bank notes.)

(12) Bokalu-ak iya mau-tu-a-n ga-bil-tza.  
    (Basque Romani)
    Gose-ak iya hil-e-a-n ga-bil-tza.  
    (Basque)
    hungry-ERG almost die-PERF-DET-LOC iPL-walk-PL
    (Romani: *bokhalo* ‘hungry’, *merav* ‘I die’)
    *We are almost dying from hunger.*

For no variety is it unknown when the mixed Romani languages came into being. Traditional, inflected Romani was still recorded in England in the 19th century, sometimes from the same people who spoke Angloromani (Smart & Crofton 1875). In the Iberian peninsula no inflected Romani was ever recorded outside Catalonia, and the earliest sources of Caló go back to the 18th, perhaps the 17th century. A Spanish author, Martin Delrio, claimed in 1608 ‘that there was a language invented by [the Gypsies] to replace their native language, which they had forgotten’ (cited in Pabanó 1915: 179, our translation). This suggests that Caló was already spoken in 1608, and that Indic inflected Romani was already lost by then, though memories of it were still extant.
The amount of Romani non-lexical elements varies from source to source. In Bright (1818) virtually all non-lexical elements are Spanish, whereas in Sales Mayo (1870) and De Luna (1951) many of these elements are Romani-derived. There seems to be a continuum between more and less Hispanicized versions, perhaps depending on the geographical area. This merits a much more detailed investigation than is possible within the scope of this chapter.

There is some debate as to the origin of the mixed Romani languages. Kenrick (1979) thinks that Indic inflection of Romani was gradually replaced by more and more English inflection. Hancock (1984) claims that it was a conscious creation in the 16th century, when Gypsies mixed with English speaking beggars, and he cites some early sources to substantiate this claim. Bakker & Van der Voort (1991) also claim a conscious and deliberate creation by adults, but they do not say when this could have happened. Boretzky & Igl (1994) claim that it is a consequence of partial language shift: there was a shift toward the language of the host country, and since there were still speakers of Romani, speakers could take over these Romani words which could then help create a language that could function as a secret code. Further research will be needed to choose between these equally plausible options.

4.7 Discussion

So far we have shown that there are mixed languages in all parts of the world and that they share a number of structural characteristics. They combine the grammatical system of one language with the lexicon from another language.

Any type of language can be involved in language intertwining. There apparently are no structural constraints that prevent the intertwining of any two languages – although the result may differ according to the typological properties of the languages involved.

The name ‘intertwining’ is chosen for the following reasons. It suggests an intricate mixture of two systems which are not necessarily of the same order, in this case lexicon and grammar. These two halves form an organic whole, and therefore one cannot remove one of the components without damaging the other component. None of the components can survive without the other. Furthermore, the term does not suggest a particular direction for the process, unlike ‘relexification’ and ‘regrammaticalization’ or ‘massive grammatical replacement’ which suggest the replacement of either the lexicon or of the grammatical system. Also, neither of the components (lexical or grammatical) in intertwining can be taken as being more important than the other. The term language intertwining is also ambiguous as to the psycholinguistic process and the historical and sociolinguistic events. These must be kept apart. Although the socio-historical facts may lead one to say that speakers of some language substituted a different grammatical system from their original
Mixed languages and language intertwining

one, the process of intertwining a lexicon and a grammar from two different sources can be defined independently from the historical directionality involved. For instance, the speakers of Ma'a are ethnically and historically Cushitic. Speaking from the perspective of their origin, they took over the Bantu grammatical system. Linguistically, however, there are arguments to claim that the Cushitic lexicon was adopted into a Bantu grammatical framework (M. Mous, in prep.). The term 'language intertwining' avoids confusion between the linguistic process and the sociohistorical facts.

The way in which intertwined languages are formed appears to be highly uniform. It is also possible to predict on the basis of the social background which languages will supply the lexicon and which one the grammar. The grammatical system (syntax, morphology, phonology) of the intertwined language is often derived from the language known best by the first generation of speakers, and from the language which it resembles most in its pronunciation (see below).

If an intertwined language is spoken by children of mothers speaking language X and fathers speaking language Y, the grammatical system will be the one from language X, the language of the mothers. This is the one mastered better, not only because of closer contacts with the mothers, but also because that language is spoken by the surrounding people – migrants tend to be men. This is true for all the cases where people of mixed ancestry speak an intertwined language. This also explains why there are regularly signs of simplification in the lexicon, in the sense that a number of fairly basic words are compounds or analytic constructions. In those cases the original term may have been unknown.

If a group who creates an in-group language as a secret language, it will always use the grammatical system of the language of the immediate surroundings. This explains why all the intertwined languages spoken by Gypsies have a Romani lexicon and the grammatical system of the language spoken in the surrounding community, and never the other way around. For some cases it is certain that the language came into being at a time when the inflected Romani language was in serious decline (e.g. Britain, Basque Country, Greece) and only used in formulas, songs and the like. Furthermore, a lexicon is remembered longer than a more intricate grammatical system, and for this reason too the decaying language is a more likely candidate for supplying the lexicon.

The phenomenon of intertwining is hardly ever reported in the literature. Nevertheless language intertwining is not as rare as it seems. As these languages are without exception in-group languages, spoken by people who also know another language, outsiders hardly get to find out about their existence. Michif, for instance, had been spoken for at least 150 years before it was first cited in a publication. And it was only after three months of fieldwork on Quechua, when Muysken accidentally discovered that his hosts spoke a different language among each other when they thought he could not overhear them: Media Lengua.
Many secret languages, argots and other in-group languages of bilingual groups in all parts of the world have a large number of lexical items taken from other languages, so that these languages in extreme cases may have a grammar and lexicon from different sources. The grammar is identical to that of the environment, but the lexicon is not.

4.8 Social conditions

Language intertwining happens under specific historical circumstances. In the first place, the group must be bilingual when language intertwining starts. Fluent knowledge of the language that provides the lexicon is not necessary, but it has to be spoken to a reasonable degree.

Second, the resulting languages are intended as in-group languages. They are not contact languages in the sense that they are intended to bridge a communication gap between speakers of different languages.

Third, the members of the group do not identify themselves as belonging to either of the groups whose languages they speak. The Ma’a have no ties at all with Cushitic groups and no special ties with the neighboring Bantu groups. The speakers of Michif identify themselves as Métis or Michif, neither as French or French Canadian nor as Cree or Indian. The speakers of the mixed languages Kröjo/Javindo and Pecoh in Indonesia identify themselves as 'Indo-Europeanen' (from Indonesian + European), not as Dutch or Indonesian.

Not in every case does the intertwined language come into being as an expression of a new identity. In most of the Gypsy cases, a new, mixed identity played only a secondary role, if at all. Although it is an aspect of Gypsy culture to avoid intensive socializing with non-Gypsies, it can be questioned whether there were other Gypsy groups around to dissociate from. In these cases a different factor was responsible for the genesis of the intertwined language: the need to be unintelligible to outsiders. The Gypsies mostly had business contacts with non-Gypsies, and that is the situation where the intertwined language would be used (Kenrick 1979). In all cases the lexicon is Romani, and the grammatical system is the language or dialect of the non-Gypsies living in the immediate surroundings. This is not chance: by making one’s in-group language sound like the language of the surrounding people, one can more easily hide the fact that one has a secret language.

The major factor, however, is that an intertwined language is an in-group language. We can actually expect similar languages to emerge between soldiers in armies protractedly residing in foreign territory, between pupils of foreign boarding schools (Smout 1988), between bilingual traders, etc., but it is unlikely that these people would consider their speech a separate language. Neither does it seem possible that they lose their ability to speak the two source languages. In short, language intertwining creates mixed languages.
twined languages differ from pidgins and creoles. It may well be a commoner type of language than people realize, even though few cases have been acknowledged and described.

Further reading
There are a few monographs relating to mixed languages: Rafferty (1982) on a Malay-Javanese mixed language, and Bakker (1992) on Michif. There are also a few comparative studies. One example is Bakker & Mous (1994). Thomason (1995) contains more detailed descriptions of these languages.