12. Mixed codes

Pieter Muysken

1. Introduction

This chapter will try to analyze some of the different types of mixed codes that have been discussed in the literature. This literature is growing rapidly and includes the edited collections of Bakker and Mous (1994) and Thomason (1996), as well as analytical studies such as Auer (1999), Bakker and Muysken (1997), Matras (2000), Muysken (2000a), Myers-Scotton (2003), Thomason (1997), and Winford (2003), and special journal issues of the *International Journal of Bilingualism* edited by Poplack and Meechan (1998), Maschler (2000), and Myers-Scotton and Jake (2000). Matras and Bakker (2003a), an important source for the present paper, have collected and contributed a number of comprehensive analytical studies, while Smith (1995) contains a detailed list of a large number of mixed languages. In addition there are a large number of monographs reporting on individual mixed languages and more detailed comparative studies. Indeed, the topic is vast, and I am certain to have neglected some crucial sources. Furthermore, the history of research of the topic remains uncharted, although Peter Bakker has been working on this for some time now.

I will preliminarily define a mixed code as a way of speaking which shows evidence of substantial amounts of morpho-syntactic and/or lexical material from at least two different languages. By itself, this definition excludes the very important domains of semantic, structural, and phonological interference (e.g. the studies in Nichols 2001, such as Dussias 2001). As a consequence, many of the phenomena involved in pidgin and creole genesis (cf. Lefebvre 1998) also fall outside the scope of this paper, since they are primarily semantic, structural, or phonological in nature, and do not necessarily involve lexical elements from different languages. Matras and Bakker (2003b: 1) limit mixed languages to those languages which (a) emerged in the setting of community bilingualism and (b) show a non-marginal etymological split in structures. Criterion (a) is uncontroversial, but criterion (b) is not very clearly formulated. Is the lexicon irrelevant? Do structures have an etymology? Finally, I will not discuss mixed codes that may emerge in bilingual child language development (Meisel 1989; Deuchar and Quay 2000), in second language development, in language attrition and language death, or in the bilingual speech of aphasic or Alzheimer patients. Discussing these would require the consideration of a host of other background assumptions irrelevant to the topic at hand.
The main questions will be treated in four separate sections:

- What are the different types of mixed codes encountered and what are their formal properties (section 2)?
- Under which social conditions did these different mixed codes emerge (section 3)?
- By which psycholinguistic processes did these codes emerge (section 4)?
- Can we account for the properties of these mixed codes and the (a) symmetries in the components of the contributing languages in terms of general models of language processing (section 5)?

It almost goes without saying that I adopt some version of the Uniformitarian hypothesis, holding that given the same circumstances of genesis, the same linguistic results will be produced in different periods of time.

2. What are the different types of mixed codes encountered and what are their formal properties?

From the perspective of the formal definition of the types of mixing encountered, we can distinguish at least thirteen different patterns:

- heavy borrowing
- slang and jargon-type relexicalization
- insertional code-mixing
- alternational code-mixing
- discourse marker switching
- congruent lexicalization
- Media Lengua-type relexification
- Mbugu-type paralexification
- restructuring and relexification
- Michif-type NP/VP splits
- Australian mixed codes
- Copper Island Aleut-type NP/VP-splits
- Mixed pidgins, trade jargons, and creoles

I will discuss these one by one in the following sections.

2.1. Heavy borrowing

This type of mixed codes results from heavy borrowing into the non-core vocabulary. This type of borrowing is subject to semantic restrictions: very specific and pragmatically salient terms are borrowed, and to category restrictions: mostly nouns and discourse markers are borrowed, and only later other elements
such as verbs and adjectives (cf. e.g. Poplack, Sankoff, and Miller 1988). A typical example from Quechua with heavy Spanish borrowing is the following (cf. also van Hout and Muysken 1994):

(1) Spanish borrowings in Quechua

ñuka *parlu-wa-da parla-gri-ni ñuka Collana-munda awilu-guna
I talk-DIM-AC talk-INC-1s I Collana-ABL grandparent-PL
parla-shka-da.
talk-NOM-AC
Chimborazu-mun-shi primero shuk pobre ri-n k’atu-na-un.
Chimborazo-DAT-REP first one poor go-3s sell-NOM-with
Chi-munda-ga ri-n, ri-n, ri-n, k’atu-na-un.
that-ABL-TO go-3s go-3s go-3s sell-NOM-with
Mana kay-bi k’atu-y pudi-sha, chay-mun k’atu-nga-bu ri-n.
not this-LO sell-INF can-SUB that-DAT sell-NOM-BEN go-3s
Chi-munda shuk *amu kaballu monta-shka tupa-sha-ga, ni-shka:
that-ABL one lord horse mount-NOM meet-SUB-TO, say-SD:

‘I will tell a story that my grandparents from Collana told. They say that first a poor man goes to Chimborazo with his goods to be sold. Thus he goes, goes, and goes, with his goods. If he cannot sell them here, goes there to sell them. Then upon meeting with a white man mounted on a horse he says: …’

The beginning of the story contains the following lexical elements from Spanish:

(2) parlu *parlo ‘story’
parla- parlar (arch.) ‘talk, tell’
awilu abuelo ‘grandparent’
primo ‘first’
pobre pobre ‘poor’
pudi- poder ‘be able’
amu amo ‘lord, white man’
kaballu caballo ‘horse’
monta- montar ‘mount’
tupa- topar ‘encounter’

In a large corpus of spoken Bolivian Quechua more ‘intimate’ or ‘advanced’ borrowing patterns can be observed. Frequent borrowed prepositions include con ‘with’, como ‘like’, and hasta ‘until’; conjunctions include cuando ‘when’, porque ‘because’, and si ‘if’; borrowed adverbs include casi ‘almost’; finally among the frequent borrowed interjections we find en fin ‘finally’, pues ‘then’,
and a ver ‘let’s see’. Although the amount and diversity of borrowings is truly remarkable in the corpus, evidence of the historical depth of the borrowing process and the intensity of contact, we should keep in mind that borrowed pronouns of any kind (personal, possessive, demonstrative, relative, interrogative) are very rare, as are borrowed articles, basic prepositions, etc. (Muysken 2000b). Many of these borrowed non-content elements may have entered Quechua as discourse markers (see section 2.5 below).

Poplack and Meechan (1998) have brought together a number of convincing papers showing how productive borrowing can be in bilingual communities. Cases studied are English/Turkish contact in Northern Cyprus, Ukrainian/English contact in Lehighton (Pennsylvania), Igbo/English mixed speech in Nigeria, Persian/English mixing among Iranian students in Ottawa, and Acadian French/English contact in Moncton (New Brunswick, Canada). These papers use the same analytical tools as, and confirm the results of, earlier work by the two editors. However, it should be kept in mind that the type of borrowing discussed here has different socio-historical properties from the type of intensive borrowing reported in example (1), which I will take to be characteristic of many such situations. I will also refer to it as ‘historical borrowing’.

*Time depth.* The studies collected in Poplack and Meechan (1998) reflect variable time depth: from longstanding (Acadian French/English in Moncton) to fairly recent contact (Persian/English in Ottawa). In contrast, the Quechua-Spanish contact leading to historical borrowing has lasted now for almost five centuries, and the first Spanish loans entered Quechua in the mid-16th century.

*Degree of bilingualism.* In the Quechua-Spanish type situation, there need not be community bilingualism (although the percentage of Spanish loan tokens will co-vary with the proportion of bilingual speakers in the community). In contrast, all the settings reported in Poplack and Meechan (1998) involve considerable degrees of bilingualism.

*Gradualness.* The introduction of Spanish items into Quechua has been a gradual process, while there is no evidence that the emergence of bilingual mixed speech of the type reported in Poplack and Meechan (1998) was a gradual process.

*Motivation and variability.* Although I am not familiar with explicit studies about this, there is a good chance that the type of bilingual borrowing described by Poplack and Meechan (1998) is variable, subject to level of informality, audience design, etc., in other words, is stylistically motivated. This holds for the type of borrowing in (1), only to a much more limited extent.

These four differences are probably the reason why there has been such resistance in the code-switching research community to accept Poplack and Meechan’s conclusions that nonce borrowing is essentially the same thing as other
kinds of borrowing. Notice that multi-fragment code-switching patterns with bilingual borrowing on these dimensions, as shown in Table 1:

**Table 1.** Historical borrowing, bilingual borrowing, and code-switching compared on four dimensions

<table>
<thead>
<tr>
<th></th>
<th>time depth</th>
<th>degree of bilingualism</th>
<th>gradualness</th>
<th>motivation and variability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>borrowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilingual</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>borrowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>code-switching</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

The formal and distributional properties of the two borrowing types are quite similar, but their embedding in the socio-historical context may be quite distinct.

2.2. Slang and jargon-type relexicalization

Similar to borrowing is what Wälchli (2005) calls relexicalization, the replacement of native vocabulary by words from one or more other languages. This process, like borrowing, is subject to category restrictions: it involves major class items, mostly nouns. However, unlike borrowing, it is not subject to semantic restrictions, and often involves the replacement of core vocabulary. The following example from Stadin Slangi Finnish, a non-standard urban variety of the language, is a typical example. Swedish words are italicized, and the inflection is Finnish.

(3) Swedish words in Stadin Slangi Finnish

*Bro* -i*en* mutsi ol-I nasta mimmi.

Brother-PL.GEN mother be-PST.3s good/nice woman

‘The brothers’ mother was a good woman’ (Wälchli 2005).

The process of relexicalization is extremely frequent in a wide variety of urban youth slangs, in jargons, secret and trade languages, etc.

There are several other characteristic features: (a) The process generally involves the replacement of large portions of the native content words, but is not categorical, and there is variation in the content words replaced. (b) Words can come from various languages, particularly in urban youth slangs. In Table 2 a few frequently cited youth slangs are listed:
Table 2. Frequently cited youth slangs

<table>
<thead>
<tr>
<th>name</th>
<th>matrix or frame</th>
<th>embedded lexicon</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheng</td>
<td>Swahili</td>
<td>English</td>
<td>Kenya</td>
</tr>
<tr>
<td>Engsh</td>
<td>English</td>
<td>Swahili</td>
<td>Kenya</td>
</tr>
<tr>
<td>Tsotsitaal-Flaaitaal (Makhudu 2002)</td>
<td>Afrikaans</td>
<td>English, Zulu, slang</td>
<td>Johannesburg</td>
</tr>
<tr>
<td>Isicamtho (Slabbert and Finlayson 2002)</td>
<td>Zulu/Nguni</td>
<td>English, Afrikaans</td>
<td>Johannesburg</td>
</tr>
<tr>
<td>Pachuco/Trilongo/Tirili</td>
<td>Spanish</td>
<td>English, slang</td>
<td>American Southwest</td>
</tr>
<tr>
<td>Tex-mex</td>
<td>Spanish</td>
<td>English, slang</td>
<td>Texas</td>
</tr>
</tbody>
</table>

A special type of relexicalization occurs when a special lexicon has been inherited from an older community language. Particularly well-known are the para-Romani languages, in which diverse (mostly European) languages can be combined with a lexicon largely based on Romani. Examples include Anglo-Romani (English grammar, Romani lexicon) and Caló (Spanish grammar, Romani lexicon). Other cases are Lekoudesch, a language of cattle traders in southern Germany with Hebrew words inserted into a German grammatical frame, and Abdal or Äynu, with Persian vocabulary in a Turkish grammatical frame (Matras and Bakker 2003b).

2.3. Insertional code-switching

Above I used the term code-switching as a generic term. However, a number of authors have argued that there are several kinds of code-switching (Sankoff, Poplack and Vanniarajan 1990). Muysken (2000a), using the cover term code-mixing, has argued that there are three main types, of which insertional code-switching is the first. Here, separate constituents from language B are inserted into a frame constituted by the rules of language A. The main restriction on this process is categorical or semantic congruence or equivalence between the inserted element and the properties of the slot into which it is inserted. A typical example is given in (4), where an (italicized) Dutch adjective + noun combination is inserted into a Turkish clause:

(4) on-dan sonra lauw water-nan yıkayınca ...
    DEM-ABL later lukewarm water-INST wash-GER

    ‘and then, while you were washing with lukewarm water’ (Backus 1996: 103).
Notice that the Turkish postposition or case marker -nan is added to the Dutch noun as if it were a Turkish noun.

It is a matter of contention whether insertional code-switching is distinct from the type of bilingual borrowing discussed in section 2.1. The majority of single constituents inserted are either single words or fixed phrases, which could be analyzed under borrowing. However, not all of them are, and in some language pairs multi-word inserted constituents abound.

2.4. Alternational code-mixing

The second type of code-switching distinguished by Muysken (2000a) is alternation. Here a chunk in language A is combined with a chunk in language B (see e.g. Poplack 1980, 1985, who adopts a slightly different perspective). The principal grammatical mechanism involved is adjunction, since the languages (here, French and Dutch) do not necessarily fit together grammatically.

(5) d’r zit me hier une femme qui n’est pas drôle
there sits me here / a woman who ...
‘Here there is a woman who is not funny.’ (Treffers-Daller 1994: 224)

Often alternation involves the switching of clause-peripheral elements such as adverbial phrases, dislocated constituents, etc.

2.5. Discourse marker switching

Possibly a special subtype of switching is discourse marker switching, the topic of a collection of articles edited by Maschler (2000). An example with the discourse marker donc (in italic) from a Shaba Swahili/French bilingual conversation (De Rooij 2000: 456) is (6):

(6) Tu-ko ba-ntu ba-moya b-a chini. donc tu-ko ba-faible. eh?
we-COP 2-man 2-DET 2-CONN low so we-COP 2-weak
‘We’re a low kind of people. So we’re weak, aren’t we?’

There is also a French adjective inserted, faible ‘weak’.

The use of a discourse marker from a different language often has a highlighting function in structuring the discourse: its very non-nativeness makes it useful to employ a foreign discourse marker. Furthermore, sometimes an element from a different language can help bring the conversation into a more informal domain.

What is important to realize is that the use of discourse markers does not obey the same directionality constraints as e.g. insertion or bilingual borrowing.
While borrowing and insertion generally involve elements from a dominant language put into a community language, this does not hold for discourse marker switching. In some cases, like the Shaba Swahili case mentioned above, indeed the discourse markers are from a dominant language; in other cases, it is the community language that provides the discourse markers, e.g. to make a community variety of the dominant language more ethnic.

Furthermore, it should be kept in mind that discourse marker switching is something very much akin to alternation, in that both involve the periphery of the clause. However, in the case of alternation, there are grammatical structures from the two languages involved, which are oddly combined, while in the case of discourse marker switching, there is the general possibility of clause-peripheral adjunction of separate elements, without internal grammatical structure.

A final point is that the frequency of discourse marker switching probably is the reason for the fact that in so many languages discourse markers have been borrowed from other languages. Frequently switched discourse markers need not become integrated into the native lexicon, but it may happen. An example of a stable non-integrated discourse marker in Dutch is German überhaupt ‘in any case’, used frequently but retaining a distinct, non-native status, which contributes to its pragmatic force.

2.6. Congruent lexicalization

The final type of code-switching distinguished in Muysken (2000a) is congruent lexicalization, the rapid back and forth switching of loose elements in a structure mostly shared by the two languages. Example (7) is from Sranan Creole / Dutch switching, in this case recorded in the Netherlands:

(7) soort bijdrage yu kan lever op het ogenblik gi a opleving type contribution you can provide at the moment for the revival of the culture?

(Bolle 1994: 78)

The elements in italic are Dutch, those in bold ambiguous between Sranan and Dutch, and the others Sranan. The Sranan word kulturu ‘culture’ is a direct though phonologically adapted borrowing from Dutch. The word soort ‘type’ has a Dutch shape, but is used as a question word the way it would be in Sranan.

Congruent lexicalization is typical of dialect/standard language switching, bilingual settings with considerable convergence, and switching between closely related language varieties.
2.7. Media Lengua-type relexification

After the phenomena discussed in 2.1–2.6, which are quite common in the languages of the world, and do not necessarily lead to stable varieties, I now turn to a number of mixed codes which are both unique and highly infrequent, and which have a much more stable existence, sometimes as the principal community language. A first such code is Media Lengua, varieties of which are spoken in several communities in Ecuador (Muysken 1981, 1996a). Media Lengua can be characterized in a nutshell as Quechua with Spanish-derived lexical roots.

\[ \text{(8) a. kuyi-buk yirba nuwabi-shka} \]
\[ \text{guinea.pig-BEN grass not.be-SD} \]
\[ \text{b. kuyi-buk k'iwa illa-shka} \]
\[ \text{c. No ha habido hierba para los cuyes} \]
\[ \text{‘There turns out to be no grass for the guinea pigs.’} \]

The vast majority of its affixes are Quechua, as is its syntax and most of the phonology. However, over 90% of the roots are Spanish; the few exceptions can be seen as borrowings from Quechua.

There are people who know Media Lengua but no Quechua, and most Media Lengua speakers at present are also fluent Spanish speakers. Media Lengua is very different, however, from Quechua/Spanish L2 interlanguage, and for a number of speakers it is or was their primary mode of communication.

Matras and Bakker (2003b) cite several languages which have properties similar to those of Media Lengua: Basters Afrikaans (Khoi-khoi structure, Afrikaans root lexicon) and Chindo (Malay/Javanese structure, Peranakan Chinese lexicon). Bakker (2003: 116) suggest that there are about 25 Media Lengua-type mixed codes, which he terms L-G languages (with a Lexicon/Grammar split), and treats as the paradigm case of mixed languages: intertwined languages.

2.8. Mbugu-type paralexification (co-existence of two lexicons)

Like Media Lengua, Ma’á or Mbugu combines roots, italicized here, from one language (Cushitic in origin) with structures from another language (Bantu), as in example (8).

\[ \text{(9) hé -lo mw -agirú é -sé -we kimwéri dilaó w -a} \]
\[ \text{16 have 1 elder 1 call PS:PF Kimweri king 1 CONN} \]
\[ \text{‘There was an elder called Kimweri, king of ’} \]
\[ \text{(Mous 2003: 9).} \]
Mous (2003) has termed the process involved in the genesis of Ma’á or Mbugu paralexification, because the original lexicon always remains available in another register.

The same holds for Callahuaya (Muysken 1996b), a semi-secret sorcerer ritual language from Bolivia, with a Quechua structure, like Media Lengua, but a lexicon with words from Puquina, in addition to various other Amerindian languages.

2.9. Restructuring and relexification

Sometimes the process of relexification goes hand in hand with considerable grammatical restructuring. Two cases in point are mixed codes from the island of Java, dating back to the Dutch colonial period, called Petjo (with mostly Malay grammar and mostly Dutch content words) and Javindo (with mostly Javanese grammar and mostly Dutch content words). In the cases of Javindo and Petjo, the ‘grammatical’ lexicon is split between elements from Dutch and from Malay or Javanese. In addition, there is some variation in this respect. An example is given in (10) from Petjo (van Rheeden 1993: 114, with spelling of the original source maintained):

(10) ja-îtoe soesah-nja feel
    yes-DEM trouble-DET much
    ‘Yes that is such a lot of trouble.’

Here the italicized Malay deictic enclitics -îtoe and -nja can be attached to Dutch words. The bold soesah is a Malay loan word in Dutch. The overall grammar is Malay, as far as we can see, but the Malay elements in Petjo are simplified with respect to the original Malay system.

2.10. Michif-type NP/VP split

The next three types of mixed codes all involve a NP/VP split. In Michif, a mixed code currently from the plains of western Canada and the adjacent parts of the US (it originated more to the east), the italicized noun phrases are French, while the verb phrase and the clause are structured with Cree principles and mostly Cree lexicon:

(11) èkwa pàstîn -am sa bouche òhi le loup
    and open-he.it his.F mouth this-OBV the.M wolf
    è-wî-otîn-åt
    COMP-want-take-he.him
    ‘and he opened his mouth and the wolf wanted to take him.’

    (Bakker 1997: 6)
2.11. Australian mixed codes

Another type of mixed code we find in aboriginal communities in northern and central Australia, notably Gurindji Kriol (GK) and Light Warlpiri (LW) (O’Shannessy 2005; Meakins and O’Shannessy, to appear). (12) is a set of examples from Warlpiri, Aboriginal English, and Light Warlpiri, which O’Shannessy (p.c.) would classify as a mixed language. Here roughly the same sentence is presented in three language varieties (data and background information from O’Shannessy, p.c.):

(12) a. Warlpiri:
Yirra-rni ka-xafl-xafl leda watiya-ngka kurdu-pardu-rlu.
put-NPST IMPF-3sg-3sg ladder tree-LO child-DIM-ERG
‘A child is putting a ladder against the tree.’

b. Aboriginal English
Dat boi bin purr-um leda on dat tri.
DEF boy PST put-TR ladder PREP DEF tree
‘The boy put a ladder against the tree.’

c. Light Warlpiri
Kurdu-pawu-ng i-m purr-um leda na watiya-wana.
child-DIM-ERG 3sg-NFUT put-TR ladder DIS tree-PERL
‘The child is putting the ladder against the tree.’

What characterizes Light Warlpiri, among other things, is a more English-type SVO order. Most verbs and verbal morphology derive from Kriol and Aboriginal English (often the difference is hard to see). Nouns can be Warlpiri and English, and nominal morphology is Warlpiri in origin (O’Shannessy 2005). The most striking feature is the innovative auxiliary system, which combines features of Kriol and Warlpiri. Notice the personal reference marker -i, derived from ‘he’, and a non-future marker -m, from Warlpiri. The English-etymon verb purr- ‘put’ has a Kriol transitive marker -um, ultimately derived from ‘him’.

The NP/VP split in Light Warlpiri, then, is the mirror image of the Michif-type. In Michif, the traditional language Cree provides the verbal system, and in Light Warlpiri the traditional language provides the nominal system. Gerrit Dimmendaal (1998) and Patrick McConvell (1998) have suggested independently from one another that the difference is due to the overall typological difference between head-marking languages like Cree and dependent-marking languages like Warlpiri.
2.12. Copper Island Aleut-type NP/VP-split

The split in Copper Island (Mednyj) Aleut is again unlike that in Michif: the finite verb morphology is Russian, with Russian pronouns in the Russian past tense (when there is no person marking on the verb):

(13) ya tibe cibu-x ukagla:ya:sa:-l
1sg.SUBJ 2sg.OBJ parcel-ABS bring-PST
'I brought you a parcel' (Golovko and Vakhtin 1990: 105).

Here the pronouns and the tense marking are Russian; the global structure is Aleut. Only a small minority of the verb roots is Russian, but the number of Russian nouns is much larger. Nominal morphology is Aleut, as well as verbal derivational morphology.

What distinguishes Mednyj Aleut from Light Warlpiri is that the verb roots themselves are largely drawn from the original language, rather than from the new language (Russian), as in the case of Light Warlpiri.

2.13. Mixed pidgins, trade jargons, and creoles

While many pidgins, jargons, and creoles have lexicons mostly derived from one language source, there is always a fair amount of admixture of foreign vocabulary, and a number of languages actually have quite mixed vocabularies. The process through which this came about has been sketched by Silverstein (1972a, b), among others, in his analysis of the emergence of Chinook Jargon. Speakers of different languages may negotiate a common mixed code when they have roughly equal power and prestige. The example given here is from Russo-norsk:

(14) kak ju wil skaffom ja drikke te, davaj på sjib tvoja ...
what you want eat and drink tea, please on ship your
‘If you want to eat and drink tea, then come on board …’
(Bakker 1995: 36/7)

Russian is bold, Norwegian italic; the rest is international nautical vocabulary.

Creoles with mixed vocabulary include Berbice Dutch (Dutch, Ijo, Arawak), Saramaccan (English, Portuguese, Fongbe, Kikongo), and Chavacano (Spanish, various Austronesian languages).
3. Under which social conditions did these different mixed codes emerge?

After enumerating these different mixed codes, the question arises under which social conditions these different mixed codes emerged. Here, Croft (2003) distinguishes five types of mixed languages:

- death by borrowing: as the number of speakers decrease and the contexts of use become more limited, more and more vocabulary and structures from a dominant language enter. The result resembles a mixed language, but not of the type that concerns us here. An example would be Asia Minor Greek, in which Greek grammatical patterns and lexicon were partly replaced by Turkish elements.
- semi-shift: speakers give up a language but relexicalize the new language with the vocabulary of the old one, in order to maintain something resembling their old ethnic identity.
- mixed marriage languages: children born in communities where many fathers speak one language, and many mothers another one, may end up speaking a linguistic variety with a grammar contributed by their mothers’ language, and a lexicon derived from their fathers’ language.
- new community languages owe their existence to the need to express a new ethnic identity.
- secret languages may be formed by relexicalizing a majority language with lexical elements drawn from an older or a minority language.

We can add several social contexts here:

- urban youth or street languages;
- ritual languages such as Callahuaya;
- trade languages created for inter-ethnic contact, such as the pidgins and jargons mentioned above.

What is striking about many of these cases is that they involve an in-group language. The exception are the trade languages, where we do not find an asymmetric syntax/lexicon_a/lexicon_b mix, but rather a symmetric lexicon_a/lexicon_b mix.
4. By which psycholinguistic processes did these codes emerge?

A number of psycholinguistic processes have been suggested in the literature to account for the emergence of mixed codes. Bakker (2003) suggests four fundamental processes, linked to four sociolinguistic scenarios:

Table 3. The four contact processes in the typology of Bakker (2003)

<table>
<thead>
<tr>
<th>scenario</th>
<th>process</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>maintenance</td>
<td>lexical mixing</td>
<td>add items from one lexicon to another one</td>
</tr>
<tr>
<td>shift</td>
<td>interference or transfer</td>
<td>bring structural elements from the old language into the new one</td>
</tr>
<tr>
<td>bilingual</td>
<td>interwining</td>
<td>matching of the lexicon from one language with the grammar of another one</td>
</tr>
<tr>
<td>creation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bilingual</td>
<td>language conversion or</td>
<td>massive grammatical restructuring under the influence of another language</td>
</tr>
<tr>
<td>resistance</td>
<td>metatypy (Ross 2001)</td>
<td></td>
</tr>
</tbody>
</table>

The variety of processes having to do with ‘lexical mixing’ (relexicalization, borrowing, lexical reorientation, etc.) need not be perceived as problematic in themselves (although the cases where it is semi-categorical are striking, of course). The process of language conversion or metatypy is not at all well understood (cf. e.g. Gumperz and Wilson 1971; Ross 2001), but falls outside the scope of this overview. The same holds for interference in a shift setting: it does not produce a mixed code as defined here. The process of interwining, which I will take here to include relexification, however, does need a special comment. There are many close links between lexicon and grammar; ideally they go together like hand and glove, even if in a modular view of our language capacities they are seen as essentially distinct cognitive components. So why split them up, and how is this mentally possible?

Two clusters of answers can be found in the literature: conscious creation and transformation of another mixing process. Thomason (2001) suggests that mixed languages must have emerged through a semi-conscious process of creation. The title of Golovko (2003) suggests “‘folk’ linguistic engineering” in this context. The word ‘conscious’ should not be taken too literally, since we are dealing with speech communities without a tradition of meta-linguistic reflection. I think it is fair to say that conscious processes of language creation will predominantly involve the lexicon (both borrowing and lexical creation) and
certain aspects of pronunciation of which speakers are sufficiently aware to manipulate (cf. Labov's 1972 work on variables, markers, and indicators). Thus we can imagine something like relexicalization and relexification as the result of conscious creation, but not the adoption of underlying grammatical patterns.

The other suggestion is the transformation of other mixing processes, part of Saussurean parole, into a grammaticalized and stable code, part of langue. An example of this may be what we find in Media Lengua. While Quechua is spoken in Bolivia, Ecuador, and Peru, completely relexified varieties are found only in Ecuador. In the other two countries incidental relexification occurs in the so-called waynos, bilingual songs, but as a special register, and not obligatorily. It may be surmised, but this cannot be proven, that the same psycholinguistic process responsible for occasional relexification in bilingual songs is used, in the specific socio-cultural settings of highland Ecuador, to produce various instances of Media Lengua as fixed speech varieties in specific communities.

Likewise, various researchers have suggested that mixed languages such as Michif, argued to be the result of 'intertwining', may be the result of code-switching. Auer (1999) develops a complex transitional scenario in which code-switching leads to 'fused lects' and then on to 'mixed languages'. Indeed, the type of Cree/French code-switching recorded by Drapeau (1995) in northern Quebec shows the same verb phrase/noun phrase asymmetries as Michif. McConvell and Meakins (2005) likewise adopt this scenario. Bakker (2003) disagrees with this general scenario, and presents seven arguments against it:

(a) The quantity of imported lexical material in code-switching (frequent but variable and generally not predominant) is very different from that in intertwining (often categorical). **Comment:** The observation is correct, but grammaticalization would lead to the categorical presence of foreign words.

(b) The semantic status of imported lexical material in code-switching (often fairly specific content words) is very different from that in intertwining (often quite generic). **Comment:** Again, this is generally correct, but frequent code-switching often involves more generic words.

(c) No documentation has been provided of a transitional stage between code-switching and intertwining. **Comment:** Again, this is generally true, but the number of documented cases of intertwining is very limited anyhow, and the Australian data referred to in section 2.11 may provide just the evidence for a transitional stages as well (currently, Felicity Meakins and Carmel O’Shannessy are completing their doctoral theses in this area).

(d) When a group speaking a mixed language moves to a new area, the grammar language is swapped for a new local language; this has not been documented for code-switching communities. **Comment:** This may be true, but this possibly has to do with the fact that we define code-switching as al-
ways non-permanent. Certain types of bilingual patterns (e.g. bilingual mixed verbs) are carried from area to area (Muysken 2000, forthcoming).

(e) Different typological properties correspond to different types of code-switching: alternation occurs mostly with flexional languages, insertion with agglutinative languages. No such dependence on typology is found in mixed languages; they follow the insertional model independently of the morphological patterns of the component languages. **Comment:** This may be true for the majority of the mixed languages, but the deviant types of Michif, Mednyi Aleut and Light Warlpiri could be analysed as being derived from an alternational pattern since they involve distinct verbal and nominal morphology.

(f) Code-switching as actually documented does not look like intertwining. Peter Bakker illustrates this with data involving Turkish mixed codes. He begins by quoting an example of Turkish/Romani bilingual usage from Istanbul (his own fieldwork data):

(15) Amen roman\(\text{s}\) \textit{konuş-uyor-uz}

\begin{verbatim}
we Romani-ADV speak-PRES-1pl
\end{verbatim}

‘We speak Romani.’

This pattern, which Bakker claims to be fairly typical for Turkish/Romani bilingual communities in Turkey and adjacent parts of the Balkan, involves the use of a fully inflected Turkish verb. It contrasts with the pattern encountered in Turkish/German and Turkish/Dutch code-switching (Backus 1996), where alien verbs are introduced by a Turkish ‘to do’ auxiliary, often \textit{yapmak}. (In addition there is less Turkish in examples such as (15) than in many code-switches reported by Backus (1996).) However, it also contrasts with what we find in ‘intertwined’ secret languages involving Turkish, where Turkish verbal inflections are added directly to alien verbs. Thus, Bakker concludes, intertwined languages involving Turkish could not have evolved from Turkish/X code-switching patterns. **Comment:** This argument is to some extent circular: the pattern illustrated in (15) and the one involving \textit{yapmak} are also to some extent grammaticalized, in the sense that they reflect an established, non-arbitrary and systematic mixing practice in a particular bilingual community. They are mixed codes in their own right. Arguably, the L-G mixed languages discussed in 2.7 may be more frequent, but this does not make them the only type. Mednyi Aleut and the Australian mixed languages are quite different, and possibly have been overlooked so far in the survey of the mixed codes of the world (cf. also Matras 2003).

(g) Code-switching and intertwining are used in different sociolinguistic circumstances. Code-switching occurs in bilingual communities in which bi-
linguals have a positive attitude towards both cultures, while intertwining is typical of ‘new’ ethnic groups. **Comment:** This is only partially correct; both tend to be in-group phenomena. Furthermore, not all code-switching communities have a positive attitude towards both languages; and some of them are in the process of becoming ‘new’ ethnic groups. It is a more gradual than absolute distinction. Also, we cannot independently reconstruct the attitudes which lead to intertwined languages.

Thus we can conclude that the arguments given by Bakker show that the standard type of intertwined languages cannot be equated with code-switching, but this does not mean that they could not be a grammaticalized and regularized result of code-switching practices. Golovko (2003) suggests that two processes: code-switching and ‘lexical reorientation’ (termed relexicalization in 2.2 above) independently of each other could lead to two different types of mixed codes: the Media Lengua-type, and the Mednyi Aleut-type. Backus (2003) argues that conventionalization of alternational code-switching (cf. section 2.4) always leads to an output in which full phrases are combined (as e.g. in (15)), never to a grammar/lexicon split. However, this does not mean that insertional code-switching could not lead to this type of ‘intertwined’ mixed language. This leaves us with two kinds of mixed languages, possibly corresponding to two kinds of code-switching.

### 5. Abstract models to account for the asymmetries in mixing patterns

Can we account for the properties of these mixed codes and the (a)symmetries in the components of the contributing languages in terms of general models of language processing? I will start out by summarizing the contribution of the different languages in the mixed codes discussed, in Table 4.
Table 4. Summary of the contributions of the different mixed codes

<table>
<thead>
<tr>
<th><strong>Asymmetrical patterns</strong></th>
<th><strong>Components</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>heavy borrowing</td>
<td>Grammar and vocabulary from A, many content words and discourse markers from B</td>
</tr>
<tr>
<td>slang and jargon-type relexicalization</td>
<td>Grammar and vocabulary from A, possibly a majority of content words and discourse markers from B</td>
</tr>
<tr>
<td>insertional code-mixing</td>
<td>Grammar and constituents from A, inserted words, phrases, and small constituents from B</td>
</tr>
<tr>
<td>Media Lenga-type relexification</td>
<td>Grammar, including affixes, from A, lexical roots from B</td>
</tr>
<tr>
<td>Mbugu-type paralexification</td>
<td>Grammar, including affixes, from A, lexical roots from A and in a different register from B</td>
</tr>
<tr>
<td>restructuring and relexification</td>
<td>Grammar mostly from A, with some patterns from neither A nor B, and lexicon from B with some words from A</td>
</tr>
<tr>
<td>Michif-type NP/VP splits</td>
<td>Verb phrases from A, noun phrases from B</td>
</tr>
<tr>
<td>Australian mixed codes</td>
<td>Noun phrases from A, verbs from B, and a compound auxiliary complex with features of both A and B</td>
</tr>
<tr>
<td>Copper Island Aleut-type NP/VP-splits</td>
<td>Noun phrases from A, verbal inflections and related pronouns from B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Symmetrical patterns</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>alternational code-mixing</td>
<td>Chunks of A and B in alternation</td>
</tr>
<tr>
<td>discourse marker switching</td>
<td>Clauses from A, with discourse markers from B, or the reverse pattern</td>
</tr>
<tr>
<td>congruent lexicalization</td>
<td>Mostly shared structures of A and B, with fairly random lexicon from both A and B</td>
</tr>
<tr>
<td>mixed pidgins, trade jargons, and creoles</td>
<td>A grammar with elements of A and B, but mostly newly formed, and lexicon from A and B</td>
</tr>
</tbody>
</table>
Leaving aside the trade pidgins and jargons, it is clear that there is an overall asymmetry in the more stable mixed codes. In particular, the distinction functional versus lexical categories and grammatical patterns versus content words play a role here as pointed out by Myers-Scotton (1993) and in a host of other publications.

The way I wish to approach the diversity of the patterns encountered, as well as the frequent occurrence of some patterns, is through a competition model in which a number of independently alternative principles may play a role in creating a mixed code. These principles would include those listed in Table 5.

Table 5. Linguistic processing principles governing the outcome of bilingual strategies

<table>
<thead>
<tr>
<th></th>
<th>Categorial hierarchy</th>
<th>Verbal elements are retained more frequently than nominal elements.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>N &lt; V</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>lex &lt; func</td>
<td>Functional elements are retained more frequently than lexical elements.</td>
</tr>
<tr>
<td></td>
<td>Functional hierarchy</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Core &lt; Non-core</td>
<td>Core vocabulary items are retained more frequently than non-core vocabulary items.</td>
</tr>
<tr>
<td></td>
<td>Lexical hierarchy</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>func ~ lex</td>
<td>Functional elements are frequently taken from the same language as lexical elements in their immediate environment.</td>
</tr>
<tr>
<td>E</td>
<td>Juxtaposition</td>
<td>In language mixing and language creation a frequent strategy is juxtaposing or adjoining chunks from different languages.</td>
</tr>
<tr>
<td>F</td>
<td>Discourse</td>
<td>Discourse markers show an entirely separate behaviour, both in terms of their distribution in the clause and the frequency with which they are borrowed, but also in terms of the directionality of borrowing.</td>
</tr>
</tbody>
</table>

Applying these principles to the mixed codes pointed to, however briefly, in this chapter, yields the overview in Table 6. In different bilingual communities, different strategies are adopted, which leads to the different outcomes described here.
Table 6. Examples of mixed codes made possible by the principles A-D in Table 5

<table>
<thead>
<tr>
<th>Asymmetrical patterns</th>
<th>Principles operant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy borrowing</td>
<td>A, B, C</td>
</tr>
<tr>
<td>Slang and jargon-type relexicalization</td>
<td>A, B</td>
</tr>
<tr>
<td>Insertional code-mixing</td>
<td>B, D</td>
</tr>
<tr>
<td>Media Lengua-type relexification</td>
<td>B</td>
</tr>
<tr>
<td>Mbugu-type paralexification</td>
<td>B</td>
</tr>
<tr>
<td>Restructuring and relexification</td>
<td>B</td>
</tr>
<tr>
<td>Michif-type NP/VP splits</td>
<td>A, D, E</td>
</tr>
<tr>
<td>Australian mixed codes</td>
<td>D, E</td>
</tr>
<tr>
<td>Copper Island Aleut-type NP/VP-splits</td>
<td>D, E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symmetrical patterns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternational code-mixing</td>
<td>D, E</td>
</tr>
<tr>
<td>Discourse marker switching</td>
<td>F</td>
</tr>
<tr>
<td>Congruent lexicalization</td>
<td></td>
</tr>
<tr>
<td>Mixed pidgins, trade jargons, and creoles</td>
<td>E</td>
</tr>
</tbody>
</table>

6. Concluding remarks

Quite obviously, the above enumeration of codes, strategies, and principles only does limited justice to the incredibly rich and varied picture of mixed codes that has emerged in the previous sections. The million dollar question that remains to be answered is: can we relate the properties of the different mixed codes to the circumstances of their genesis and use? Various authors have attempted to answer this question (Croft 2003; Matras 2003), but my impression is that the answers are not yet complete. I think it is important to start exploring a richer variation in mixed codes, beyond the grammar/lexicon prototype, before we can answer this question in full.
List of abbreviations used

Is first, second, etc. person singular
ABL ablative
ABS absolutive
AC accusative
BEN benefactive/purposive
COMP complementizer
CONN connective
COP copula
DAT dative
DEF definite
DEM demonstrative
DET determiner
DIM diminutive
DIS disjunctive
ERG ergative
F feminine
GEN genitive
GER gerundial
IMPF imperfective
INF inchoative
INST infinitive
LO locative
NOM nominalizer
M masculine
NFUT non-future
NPST non-past
OBJ object
OBV obviative
PERL perative
PL plural
PREP preposition
PS:PF passive, perfective
PST past
REP reportative
SD sudden discovery evidential tense
SUB (adverbial) subordinator
SUBJ subject
TO topic
TR transitivizer
References

Arends, Jacques, Pieter Muysken and Norval Smith (eds.)

Auer, Peter

Backus, Ad
1996 *Two in one. Bilingual Speech of Turkish Immigrants in the Netherlands.* Tilburg, Tilburg University Press.

Bakker, Peter

Bakker, Peter and Maarten Mous (eds.)
1994 *Materials on Mixed Languages.* Amsterdam: IFOTT.

Bakker, Peter and Pieter Muysken

Bolle, Tiba

Croft, William

Deuchar, Margaret and Suzanne Quay

Dimmendaal, Gerrit

Drapeau, L.

Dussias, Paola

Golovko, Evgeniy V.
Golovko, Evgeniy V. and Nicolai B. Vakhtin

Gumperz, John and Robert Wilson

Hout, Roeland van and Pieter Muysken

Labov, William

Lefebvre, Claire

McConvell, Patrick

McConvell, Patrick and Felicity Meakins

Makhudu, K.D.P.

Maschler, Yael (ed.)

Matras, Yaron

Matras, Yaron

Matras, Yaron and Peter Bakker (eds.)

Matras, Yaron and Peter Bakker

Meakins, F. and C. O’Shannessy


Meisel, Jürgen

Mesthrie, Rajend (ed.)
Milroy, Lesley and Pieter Muysken (eds.)

Mous, Maarten
2003 The Making of a Mixed Language: The Case of Ma’aa/Mbugu. Amsterdam: Benjamins

Muysken, Pieter
forthc. From Colombo to Athens: Linguistic Areas for Bilingual Patterns.

Myers-Scotton, Carol

Myers-Scotton, Carol and Janice L. Jake (eds.)

Nicol, Janet L. (ed.)

O’Shannessy, Carmel

Poplack, Shana

Poplack, Shana, David Sankoff, and Christopher Miller
1988 The social correlates and linguistic processes of lexical borrowing and assimilation, Linguistics 26: 47–104.

Poplack, Shana and Marjory Meechan (eds.)
Rheeden, Hadewych van  

Rooij, Vincent de  

Ross, Malcolm  

Sankoff, David, Shana Poplack, and Swathik Vanniarajan  

Silverstein, Michael  

Slabbert, Sarah and Robert Finlayson  

Smith, Norval  

Thomason, Sarah G. (ed.)  

Thomason, Sarah G.  

Thomason, Sarah G.  

Treffers-Daller, Jeanine  

Walchli, Bernhard  

Winford, Donald  

**Recommended readings**

Matras and Bakker (2003a) contains a number of analytical studies, while Smith (1995) is a comprehensive list of a large number of mixed languages. Case studies are presented in Bakker and Mous (1994) and Thomason (1996). Myers-Scotton (2003), Thomason (1997) and Winford (2003) contain overview chapters on mixed languages (cf. also the introductory paragraph of the present article).