

The Mixed Language Debate



Trends in Linguistics

Studies and Monographs 145

Editors

Walter Bisang

(main editor for this volume)

Hans Henrich Hock

Werner Winter

Mouton de Gruyter
Berlin · New York

The Mixed Language Debate

Theoretical and Empirical Advances

Edited by

Yaron Matras

Peter Bakker

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

⊗ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

ISBN 3-11-017776-5

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

© Copyright 2003 by Walter de Gruyter GmbH & Co. KG, D-10785 Berlin
All rights reserved, including those of translation into foreign languages. No part of this
book may be reproduced or transmitted in any form or by any means, electronic or mechan-
ical, including photocopy, recording or any information storage and retrieval system, with-
out permission in writing from the publisher.
Cover design: Christopher Schneider, Berlin.
Printed in Germany.

Contents

The study of mixed languages	1
<i>Yaron Matras and Peter Bakker</i>	
Social factors and linguistic processes in the emergence of stable mixed languages	21
<i>Sarah G. Thomason</i>	
Mixed languages and acts of identity: An evolutionary approach	41
<i>William Croft</i>	
What lies beneath: Split (mixed) languages as contact phenomena	73
<i>Carol Myers-Scotton</i>	
Mixed languages as autonomous systems	107
<i>Peter Bakker</i>	
Mixed languages: Re-examining the structural prototype	151
<i>Yaron Matras</i>	
Language contact and group identity: The role of “folk” linguistic engineering	177
<i>Evgeniy V. Golovko</i>	
The linguistic properties of lexical manipulation and its relevance for Ma’á	209
<i>Maarten Mous</i>	
Can a mixed language be conventionalized alternational codeswitching?	237
<i>Ad Backus</i>	
Not quite the right mixture: Chamorro and Malti as candidates for the status of mixed language	271
<i>Thomas Stolz</i>	

Index of subjects 317

Index of authors 319

Index of languages and language groups 323

The study of mixed languages

Yaron Matras and Peter Bakker

1. Introduction

One seldom encounters a language that is not, in some way, mixed, in the sense that it has been influenced at some point in its history by another language, and employs some structure or form that derives from that language. In the discussion context on Mixed Languages, we are concerned, however, with varieties that emerged in situations of community bilingualism, and whose structures show an etymological split that is not marginal, but dominant, so that it is difficult to define the variety's linguistic parentage as involving just one ancestor language. This definition – a bilingual mixture, with split ancestry – is the one most commonly applied in the literature to Mixed Languages (see Bakker and Mous 1994; Thomason 1995, 1997d). In this chapter we offer an introduction to the study of Mixed Languages (MLs). We begin by reviewing types of ML, and illustrating them with examples of the most common varieties discussed in the literature so far. We then survey theoretical hypotheses concerning the formation, emergence and theoretical relevance of MLs to language contact studies and historical linguistics. Finally, we review the issues that occupy the centre stage position in the current Mixed Language debate, and in this volume in particular.

2. Types of Mixed Language

Even the minimal definition which we have just mentioned does not cover all languages referred to in the literature as 'mixed' in a straightforward and unambiguous manner. As discussed below, some varieties that are regarded as MLs did not emerge in a situation of fluent bilingualism. In relation to some others, the issue of languageness may be controversial (cf. Thomason 1997a, 1997d), as they may be registers that are limited to certain contexts or styles. In some of the languages, the structural admixture may not necessarily prove an obstacle to genetic classification. We begin with the set of MLs that match the definition most closely: native languages of a community of speakers, which emerged in situations of full bilingualism, and which defy conventional notions of genetic classification. A core methodological problem in the study of MLs is, that this set is very small.

Michif is in fact the only language for which we have evidence that it is used as a native language, independently of speakers’ knowledge of any of its source or ancestor languages. Michif is spoken by several hundred Métis in scattered communities in the Canadian prairie provinces and in adjacent areas in the United States. The Métis are descendants of French-Canadian fur traders and Amerindian women. The men had trekked west and married the local women, who were mostly speakers of Plains Cree as a first or second language. Their children grew up with both languages, Cree and French, but were then separated from both French Canadian and Amerindian culture. They developed their own identity, first as *La Nouvelle Nation*, later called *Métis*. The earlier French word *Mitif* is the source for the name of the language, which is also the in-group name of the Métis as a people. The language combines verbs from Plains Cree and noun phrases from French (see Table 1).

The combination of structures from both etymological sources can be illustrated by the following example – a European fairy tale, part of the Métis cultural inheritance (Bakker, field data). French-derived elements are italicised:

Table 1. Structural profile of Michif (after Bakker 1997)

Source language	Cree	French
Lexicon	Verbs	
	Few nouns	Nouns
		Adjectives
Inflection	Verb inflection	Noun inflection
Function words	Personal pronouns	
	Possessives with Cree nouns	Possessives
	Demonstratives	
		Definite articles
		Indefinite articles
	Indefinite pronouns	
	Some negators	Most negators
	Most adverbs	Few adverbs
Syntax		Numerals
	VP word order	
		NP word order
	Some subordinations	Some subordinations
	Conjunctions	Conjunctions

- (1) kayâs *une fille* *La Cendrieuse* kî-isinihkâ-sô-w.
 long-time-ago a-F girl The Cinderella PAST-name-REFL-3
avec o-mâmâ-wa kî-wîki-w *puis trois ses soeur(s)*
 with POSS-mother-OBV PAST-live-3 and three POSS-PL sister
La Cendrieuse mâka *tout* kî-piskeyiht-am *tout*
 The Cinderella however all PAST-clean-it all
La maison, le plancher kî-kisîpêkin-am
 the house the floor PAST-wash.by.hand-it
 ‘A long time ago there was a girl called Cinderella. She was living with
 her mother and her three sisters. Cinderella, however, cleaned every-
 thing. She washed the house, the floor.’

Only very few Michif speakers today are able to speak Cree and French, though Michif speakers are all fluent in English, and English has taken over the function of the principal community language. The fact that Michif is largely isolated from its two source languages makes it what we might call a ‘plain’ mixed language – and in fact, the only attested example of a ‘plain’ ML that has lost contact with both its ancestral languages.

Mednyj Aleut or Copper Island Aleut (CIA) is spoken by a handful of elderly people from Copper Island, one of the Commander Islands, from which they have been relocated. Its speakers identify as Aleuts, and they call their language Aleut, even though they are aware of the fact that it differs considerably from other Aleut varieties. The speakers are also referred to as ‘creoles’, as their group emerged as a result of mixed marriages between Russian fur traders and Aleut women in the early 1800s. Like Michif, Mednyj Aleut is the linguistic outcome of mixed households giving rise to a new ethnic identity, albeit of a rather small group of people. The population of Copper Island never reached more than 500 at any one time, and maybe fewer than a thousand people have ever spoken the language in the course of its history. Unlike Michif, however, there is no evidence for a complete breakaway of all linguistic ties to the source languages, and so no evidence for an unambiguously independent variety. While knowledge of Aleut appears to be declining in the community, Russian is gaining ground, and all speakers are fluent in it. Consequently, it is sometimes difficult to ascertain which Russian structures are inherently part of Mednyj Aleut, and which are on-the-spot codeswitches into Russian. Table 2 illustrates the **conventionalised** structural profile of Mednyj Aleut, by etymological source (cf. Golovko 1994; Thomason 1997c):

The following are examples of individual utterances in Mednyj Aleut, from Vakhtin (1998). Russian-derived elements are italicised:

Table 2. Structural profile of Mednyj Aleut (after Golovko 1994)

Source language	Aleut	Russian
Lexicon	90% of lexicon	Some lexicon
Inflection	Noun inflection	Verb inflection
Function words	Personal pronouns (object) Demonstratives Indefinite pronouns Some adverbs Numerals	Personal pronouns (subject) (possessives) Negators Some adverbs (Numerals)
Syntax	NP word order Some subordinations	VP word order Conjunctions

- (2) *Vchira angalim guzuu qaka-yaa-l, tin ni-qakaa-chaa-l.*
 yesterday day full dry-CAUS-3PAST 3SG.OBJ NEG-dry-CAUS-3
 ‘Yesterday it was drying all day, (but) never got dry.’

- (3) *Stiklaa-ḡ siḡa-ḡtaa-y-it davnu.*
 glass-CASE break-RESULT-3SG long-ago
 ‘The glass has been broken for a long time.’

- (4) *Ya bud ivo hayaa-t’ ukushka-ḡ haksii-t.*
 I will him ask-INF window-CASE open-INF
 ‘I will ask him to open the window.’

Media Lengua has a Quechua-derived grammatical system and a Spanish-derived lexicon. It is spoken both as a first and as a second language by Indian peasants, weavers and construction workers and their families, in a semi-rural area near the town of San Miguel de Salcedo in Ecuador. Its use is restricted to the community, and many speakers know both source languages, Quechua, which they use with Indians higher up in the mountains, and Spanish, which is used with non-Indians. Thus, although Media Lengua, like Mednyj Aleut,

is regarded as a conventionalised mixture (see Table 3), it is not entirely isolated from its source languages. Nonetheless, Media Lengua is not an on-the-spot mixture by Quechua-Spanish bilinguals. Spanish items are usually regularised and adapted completely to Quechua phonology (e.g. Sp. *venir* > ML. *bini* ‘come’, Sp. *decir* > ML *dizi* > *zi* ‘say’). Some grammatical formations, such as reduplication, are common in Media Lengua but unknown in Quechua. Most outstanding is perhaps the Media Lengua system of pronouns, which combines Spanish singular forms such as *yo/miu*, *bos/usti*, *el/listi*, with hybrid plurals, marked by the Quechua plural suffix *-kuna* (*miukuna*, *boskuna*, *elkuna*, etc.). The bulk of the vocabulary is Spanish-derived, which distinguishes Media Lengua from Quechua dialects with a strong Spanish influence (the latter tend to have no more than 40 per cent Spanish-derived vocabulary).

Muysken (1997) assumes that the language emerged between 1920 and 1940, possibly as a result of the construction of the Quito-Guarjaquil railway, which made the region more accessible and led to urban expansion and labour migration. Male adolescents are assumed to have been instrumental in the creation of the language. Unlike Michif and Mednyj Aleut, speakers of Media Lengua do not constitute a separate ethnic group, although they do occupy an intermediate geographical and social position between the Quechua speakers of the highlands, and the Spanish speakers of the valley.

Example (5) shows a sentence sample (from Muysken 1997), with Spanish-derived items in italics.

Table 3. Structural profile of Media Lengua (after Muysken 1981, 1997)

Source language	Quechua	Spanish
Lexicon		85% of lexicon
Inflection	Verb and noun inflection	
Function words	(Plural pronouns)	Pronouns Demonstratives Indefinite pronouns Negators Adverbs
	Conjunctions	Conjunctions Numerals
Syntax	Word order Subordinations	

- (5) *Isi-ga asi nustru barrio-ga asi kostumbri-n abla-na.*
 this-TOP thus our community-TOP thus accustomed-3 talk-NOM
 'In our community we are accustomed to talking this way.'

Comparable, perhaps, with these cases of conventionalised mixtures is **Chindo** or so-called **Peranakan Chinese**. The Peranakans of Indonesia are descendants of Chinese traders and Indonesian women. They see themselves as a separate ethnic group, and are regarded as such by outsiders (including Javanese and Chinese) as well. The word Peranakan means among others 'local-born person' or 'persons of mixed ancestry', more specifically of Chinese descent, but it is also used to refer to the Arabs in Java, and in the past it was used for mixed Indian-Malay people. The Peranakan Chinese speak various languages, including Low Javanese and often Indonesian. In group-internal communication, they use a mixed language that combines the grammatical system of Javanese with the lexicon of Malay (Wolff 1982; Dreyfuss and Oka 1979). In this speech form, according to studies by Dreyfuss and Oka, 88 per cent of the grammatical affixes are Javanese, but only 14 per cent of the noun roots and 4 per cent of the verb roots. Function words, adjectives and adverbs show a roughly equal distribution between Javanese and Malay. The Peranakans describe their language as *bahasa campuran* 'mixed language'.

So far we have seen two types of ML: A 'plain' ML, Michif, which is the native language of the community, and which is only historically, but not synchronically, connected to its source languages; and 'conventionalised' MLs, Mednyj Aleut and Media Lengua, as well as Peranakan, which are not the only community languages, but are spoken alongside at least one of the source languages. Common to all three is the fact that they emerged in situations of full bilingualism. In many MLs, however, bilingualism during the emergence phase is a matter of degree, and not at all unambiguous. **Ma'á**, for instance, is the language of a group that refers to itself by this name, and that lives in and around Usambara in the Eastern Province of Tanzania (see Mous, this volume). They are also referred to by their neighbours as Mbugu. As a group-internal form of speech, the Ma'á incorporate a special lexical inventory into their Bantu language Mbugu. This vocabulary is derived largely from Southern Cushitic, and partly from other sources. Although some authors have viewed Ma'á as the product of gradual borrowing of Bantu grammar into a Cushitic language (see Thomason 1997b), it appears rather to represent the conscious retention, over many generations, of Cushitic vocabulary, and addition into a special lexicon of vocabulary from other sources (see Mous, this volume). The Ma'á special vocabulary is thus the remainder of a language that has been abandoned, as well as the product of lexical borrowing and lexical

creations. Like the other MLs discussed above, Ma'á is also a symbol of ethnic separateness. But it is not the primary language of the community; rather, it figures alongside 'normal' Mbugu, as a special group-internal register characterised by the insertion of lexical (and pronominal) items from the special vocabulary reservoir (see Table 4).

Ma'á is in some respects a borderline ML type: In terms of function, it is a highly conventionalised register employed at the discourse level, and thus comparable to some extent to Mednyj Aleut and Media Lengua. On the other hand, from a structural viewpoint it consists primarily of a lexical reservoir, and diachronically it represents the selective retention of vocabulary, following language shift, with few traces of grammar (function words, such as pronouns and demonstratives), and no traces of inflection.

Selective retention of vocabulary after language shift is a feature that is common in another type of mixed languages, which we might term **inherited special lexicons**. The varieties known as **Para-Romani** are frequently cited as examples. Spoken mainly in various Romani or Gypsy communities across western Europe (Scandinavia, Britain, the Basque Country, Spain, Catalonia; but also in some communities in Greece, Turkey, and Hungary), they represent the selective retention of vocabulary in communities in which a shift from Romani to the surrounding majority language has taken place. The retained structures are primarily lexical content words, but some function words may also be retained, e.g. numerals, indefinites, some prepositions and location adverbs, and some pronouns (see discussion in Matras 2002: ch. 10; Matras 1998; Boretzky and Igla 1994). Although quoted samples of Para-Romani varieties often show a maximum number of Romani-derived lexical insertions

Table 4. Structural profile of Ma'á (after Mous 1994)

Source language	Cushitic	Mbugu
Lexicon	Most basic lexicon; some of unknown origin, possibly loans and camouflaged formations	Some lexicon
Inflection		All inflection
Function words	Pronouns Demonstratives	 Other function words
Syntax		All syntax

into a grammatical sentence framework based on the respective majority language, there is no actual documentation of spontaneous stretches of speech in a Romani mixed variety. Rather, Para-Romani today is best described as a style of speech, consisting of occasional lexical insertions into utterances in the majority or dominant (non-Romani) language. Speakers with a knowledge of a fairly large Romani-derived vocabulary will, however, be in a position to produce, on demand, sentences showing a maximum density of such insertions, and it is possible, though not obvious, that at an earlier stage in the history of Para-Romani such intensified mixing may have been the unmarked choice in community-internal communication. The example shows a phrase in Caló, or Spanish-based Para-Romani, as documented by Leigh (1998):

- (6) *Mansa* *camel-o* *tuque*.
 I (< Rom. 1SG.INSTR) love (Rom. love.3SG)-1SG you (< Rom. 2SG.DAT)
 ‘I love you.’

Another case of an inherited special lexicon is **Lekoudesch**, the secret language of Jewish cattle-traders in Germany and adjoining regions (Netherlands, Alsace, Switzerland) (see discussion in Matras 2000). Its base is the Judeo-German dialect, with dense insertions from Ashkenazic Hebrew, based on the Hebrew vocabulary that was transmitted in Jewish communities through the religious scriptures. The example, from the village of Rexingen near the Black Forest region (Matras, field data), shows Hebrew-derived items in italics:

- (7) *Lou dibr-a*, *der Guj schäff-t!*
 NEG speak.INF the man sit.3SG
 ‘Don’t talk, there’s a man present!’

Although these insertions were understood by all members of the Jewish community, they were used strictly in the context of group-internal communication among the traders, as a secret language. The Lekoudesch vocabulary in fact includes a limited number of lexical creations based on Hebrew lexical items.

A somewhat comparable case is **Callahuaya**, a secret ritual language used by itinerant healers in Bolivia (Muysken 1994, 1997b). The Callahuaya village is a ‘Quechua’ speaking enclave in Aymara territory, and dialectal evidence suggests a mixed origin of the Quechua elements, from the Cuzco and Ayacucho regions in Peru. The secret variety involves insertion of lexicon that derives from Puquina, an extinct and unrelated Amerindian language of the area that became extinct in the early 1900s. Yet another case of an inherited

special lexicon is possibly the speech of the **Abdal** or **Äynu** of Anatolia, central Asia and Chinese Turkestan (Xinjiang) (Ladstätter and Tietze 1994). This secret, group-internal variety consists of the insertion of special vocabulary primarily of Persian origin into the everyday Turkic language of the community (Turkish, Azeri or Uyghur). Like the Jewish cattle-traders, and the Gypsies of Europe, the Abdal tend to specialise in service-providing economies, which often involve travel, and it is assumed that the bulk of the vocabulary was adopted from Persian as the major trade language in the region, or perhaps as a result of mixed marriages between Persian traders and local women. In addition to Persian words, there are also words of unknown origin, and items shared with the secret lexicons of various peripatetic populations of Central Asia, among them Persian-based camouflaged or cryptolectal formations.

This places Abdal in close functional and structural proximity to another class of MLs, which we might simply call **special lexicons of (mixed) foreign origin**. This class is wider than that of the inherited special lexicons, in that we have no evidence that transmission from one language that was once spoken or written in the community is responsible for their emergence and composition. Rather, we are dealing here with secret lexicons, usually of peripatetic populations, that rely heavily on foreign vocabulary, rather than on internal cryptolectal formations (although typically both strategies are found alongside one another). Some secret lexicons used by the Jenisch travelling populations of southwestern Germany, for example, derive their vocabulary largely from Hebrew, or from Romani, with various degrees of admixture of German Rotwelsch (cryptolectal) structures. This can be explained through population admixture, as well as through contacts with ethnically distinct travelling populations and the diffusion of the latter's own group-internal vocabulary into the speech of neighbouring populations. Polari, the Romance-based vocabulary of British entertainers, fits this profile, as do some of the Anatolian secret lexicons called Karachi or Mitrip, which incorporate Kurdish and Domari elements.

Secret lexicons of foreign origin are indeed a widespread phenomenon that is closely akin to the creation of manipulated speech (lexical manipulation or lexical camouflage) for use in individual utterances. But they share several properties with the MLs of the plain and conventionalised type: They show mixed parentage, with lexicon from one language tending to combine with inflectional grammar from another; and they are markers of group identity. On the other hand, they are at the lower end of a functional continuum of speech varieties ranging from everyday native community languages, through conventionalised registers, and on to the occasional insertion of individual lexical items from a foreign lexical reservoir into individual utterances. This

position on the functional continuum often coincides, in the diachronic dimension, with rather irregular contact with the donor language of the special lexicon, further hedging the applicability of the term ‘bilingual mixtures’ (cf. Thomason 1995, 1997a).

Above we mentioned as part of the tentative definition of MLs the fact that they are of ambiguous, or perhaps of hybrid parentage, and so cannot be classified genetically in a straightforward way. In varieties that are on the low side of the functionality or languageness continuum, this of course applies only to the idealised mixed utterance, which in actual conversation may be rather infrequent. Thus, the everyday speech of the English Gypsies (Romanies) is unambiguously English, into which occasional Romani-derived insertions are made (though it may be possible that such mixing had once been the unmarked choice in the community). In such cases, any assumption of language-genetic or classificatory ambiguity would relate strictly to the mixed portion of an utterance, or to the mixed portion of the discourse. We now turn to a structural continuum of mixtures, and to types of ML where the case of a structural mixture resulting in genetic ambiguity is even more difficult to make.

The first type might be characterized as **radical re-structuring** of a language, as a result of mixture. **Javindo** is the name given to a mixed language on Java, Indonesia (De Gruiter 1994a, 1994b). The name combines the words Java and Indo, the Dutch word for a person of mixed Indonesian and Dutch descent. Another, more commonly used name for the group and the language is apparently *Krontjong*, which however has pejorative connotations. A similar language is **Petjo** (van Rheedeen 1994), the speech of the Indos with Malay speaking ancestors. Javindo speakers are descendants of Dutch men and Javanese women. Their group-internal speech, a marker of their separate ethnic identity, consists primarily of Dutch lexical morphemes, re-arranged in accordance with a Javanese sentential blueprint, and with a simplified morphology, combining Lower Javanese, Malay and Dutch derivational and inflectional morphology. The circumstances of the emergence of this variety, and the symbolic functions that it has (flagging group identity), are similar to those of plain or conventionalised MLs. Here are two examples from Javindo (De Gruiter 1994b: 66, 115):

- (8) Kopen-nja itu maarraak lho sij!
 buy-the that in-a-spreed DM she
 ‘She just buys all kinds of things!’
- (9) Lho Miel, seggen-nja si Fred siek. Apa al beter nu?
 DM name say-the mister Fred ill. How already better now
 ‘Listen Miel, it is said that Fred is ill. Has he become healthy now?’

There are languages that justify a type of ML comprising **mixed creoles**. An example is **Chavacano**, also known as Zamboangueño or Mindanao Creole Spanish (Forman 1972), widely spoken on the island of Mindanao in the Philippines (Frake 1971). The source of most lexical and grammatical morphemes is Spanish Creole, but the syntactic framework and relations between categories (for instance animacy and definiteness, aspect) tend to follow an Austronesian blueprint. Austronesian grammatical morphemes include plural markers, animate definite articles, past-tense existential verbs, and transitive derivation markers, while the pronoun system is mixed: The singular pronouns are all based on Spanish Creole. In the plural, Austronesian forms seem to co-exist with Spanish Creole forms (e.g. second person *kamo* alongside *ustedes*), but in the first person plural, Austronesian forms, which mark the exclusive/inclusive distinction (*kamí*, *kitá*), seem to be preferred.

Another case is **Berbice Dutch**, a moribund language of the interior of British Guyana (Kouwenberg 1994). The last speakers were apparently Arawak Indians who had been acculturated into a community of former slaves from Dutch plantations. It is one of few creoles based on Dutch, though much of its vocabulary, as well as a number of grammatical affixes, including bound tense markers, derive from the West African language Eastern Ijo. Both the split in lexifier languages, and the presence of inherited bound morphology from an African language, make this an exceptional creole.

Finally, we turn to a class of languages that have occasionally been treated in the literature as mixed, although they do not typically satisfy two of the more crucial criteria we mentioned above: they do not seem to resist attempts at genetic classification, nor can they be said to have emerged in situations of bilingualism, that is, to be substantially different from earlier stages of the language before the incorporation of structures from a second source language. Since continuity of transmission is not a controversial issue in the language's diachrony, it would seem appropriate, therefore, to regard these as cases of **extremely heavy borrowing**. The difficulty lies in sorting out what, in borrowing, is considered 'heavy', and what is 'extremely heavy'. Both notions take for granted that there are some patterns of borrowing that are quantitatively as well as qualitatively rare, and that some are even more rarely attested.

The most widespread feature to promote the ML candidacy of such languages is the license for frequent insertion of phrases from a second language. Examples are Persian insertions in administrative and scholarly Ottoman Turkish, Hebrew insertions in scholarly Yiddish, Italian and English insertions in Maltese, or Spanish insertions into older forms of Chamorro. All of these cases, however, differ quite fundamentally from plain and conventionalised MLs in that the insertions are open to choice, and so they can not be said to be conventionalised in the same sense as the combination of structures at-

tested in MLs, and they rely on the speaker's or writer's active knowledge of two linguistic systems. While the element of optional choice is common also in inherited lexicons, active knowledge of the source language is less common in in-group lexicons. Arguably, it is as a result of these frequent insertions that the languages in question end up borrowing a considerable amount of vocabulary and grammatical structure from their contact language. But the results differ considerably (see discussion of Maltese and Chamorro in Stolz, this volume, and of Domari in Matras, this volume). It seems, then, that 'heavy borrowing' in our present context can only be defined against two borderlines: First, against the borderline of more conventional borrowing. Second, against a clearer notion of which density of structures, inherited from the same source, is needed in order to satisfy the condition of a genetically unambiguous or classifiable language. This question is addressed by some of the contributions to this volume.

3. Theories on the emergence of mixed languages

A central criterion in identifying MLs is the degree to which they show indisputable genetic affiliation, and the degree to which they constitute 'new languages', that are not merely a continuation of their ancestor language(s). These two criteria are inter-connected, in that it is the contribution of two parent languages that interferes with continuation in the lineage, thus forming a new language, one that is more radically different from its ancestor than any conventional descendant language. Thomason and Kaufman (1988) had taken the view that some MLs owe their existence to a gradual development by which grammatical structures are adopted from a contact language, until finally none of the two source components is overwhelmingly dominant, and so the product is a language that can not be affiliated, with any clear preference, to any individual parentage. As examples they cite Mednyj Aleut, Ma'á and Anglo-romani. MLs, by this account, can arise through essentially the same processes as lexical and grammatical borrowing, the difference being in the quantity (and arguably also the quality, that is the category affiliation) of borrowed structures that are ultimately accumulated, and which then blur genetic affiliation. The subsequent emergence of a discussion context devoted to MLs has, however, tended to emphasise at least two factors that do not figure in Thomason and Kaufman's discussion. The first is the abrupt emergence of MLs (see Thomason 1995, 1997a). The second is the question whether the processes involved in ML emergence are unique, and quite distinct from gradual borrowing.

A suggestion in the latter direction had already been made by Muysken

(1981), who termed the process responsible for the creation of Media Lengua ‘relexification’. The notion behind the term was that, unlike cases of borrowing, the replacement of the lexicon is wholesale. Moreover, it is based on the substitution of inherited lexical items by one-to-one structural and functional equivalents. This too indicates a process that is quite distinct from either the mere combination of two languages, or the replication of structures from one language in the framework of the other. For Arvanitika, a case of an inherited special lexicon (Albanian lexicon in Greek), Sasse (1992) assumes relexification after a shift has taken place, with younger speakers having only partial access to the vocabulary of their ancestral language. Similar scenarios have been proposed by Boretzky and Igla (1994) and by Matras (1998, 2002) for the emergence of Para-Romani varieties.

A different approach is taken by Bakker (1997: 192–213) with his model of ‘language intertwining’ (see also Bakker and Muysken 1995, and Bakker and Mous 1994b). Bakker questions the substitution of either lexical or grammatical components of a language. Instead he attributes the emergence of mixed varieties to one single and pre-determined process involving mixed populations, by which the grammar (bound morphemes and some free grammatical morphemes) of one language, typically that spoken by native women (or, in the case of nomadic populations, the surrounding majority language), combines with the lexicon of another, usually a colonial language spoken by men (or in the case of nomadic populations, the ancestral language). The intertwining model views both contributing languages as hierarchically equivalent. The process by which they are combined is viewed as a rapid process, and quite often one that is intrinsically connected to the process of formation of a new ethnic identity. MLs in general are regarded in the intertwining model as markers of distinct identity.

It is in this connection that the hypothesis of a conscious creation of mixed languages has been formulated. Golovko (1994: 117) assumes that language mixture in Mednyj Aleut began as a game among adult Aleuts learning Russian, was then used as a secret code, and later conventionalised. A similar idea is expressed by Thomason (1995: 29) with regard to the general phenomenon of abruptly emerging mixed languages. Bakker (1997: 213) assumes that intertwined languages are created “more or less consciously”, and Mous (1994, also this volume) argues for deliberate creation of Ma’á as a register of Mbugu by speakers aiming to set themselves apart from Bantu speakers. Matras (2000) has similarly suggested that the substitution of lexicon and free grammatical morphemes must be seen as a conscious process of either ‘lexical re-orientation’ (toward a new target language), or ‘selective replication’ (of elements of an ancestral language).

Since MLs emerge in bilingual settings, some authors have regarded them as the outcome of a sedimentation process of language mixing at the utterance level. Myers-Scotton (1998) speaks of a ‘matrix language turnover’, a process by which a new language is adopted as the matrix language responsible for grammatical morphemes (grammaticiser language), while the ‘old’ community language serves strictly as the embedded language, supplying lexical morphemes (lexifier language). The result, the lexicon-grammar split typically observed in MLs, resembles a consistent pattern of mixing whereby one language serves as a matrix, the other as the embedded variety. Auer (1999) has interpreted MLs as the far point on a continuum involving the acceptance of mixing first at the discourse, then at the utterance level, as an unmarked choice.

At this point let us summarize some of the principal issues and dilemmas facing the interpretation of MLs and their structures and emergence histories. On the historical-linguistic dimension, that of language genesis, the question arises whether MLs can be viewed as cases of language maintenance (with absorption of contact language material) or, respectively, language shift (with retention of substrate language material). An alternative viewpoint sees MLs as new languages, only partly related to either of their linguistic ancestors. Either viewpoint touches on the question of which structures must be relied upon in order to postulate linguistic-genetic affiliation, in view of the fact that MLs tend to show a split between the source language of much of the grammar, and that of much of the lexicon.

Included in the issue of genesis is the question whether MLs arise through abrupt or gradual processes. For some MLs, such as Michif and Mednyj Aleut, abrupt processes of emergence are attested (although no documentation of the earlier stages exists, the social circumstances that gave rise to contact are known). For others, such as Para-Romani varieties or Ma’á, we know of prolonged existence, and of some erosion over the last few generations in the structures of the lexifier language (Romani and Cushitic, respectively), but we have no clear indication of the initial process of emergence, which remains open to controversy.

There is rather wide agreement about the kinds of social constellations that lead to the emergence of MLs: They typically arise either in communities with mixed households accompanying the formation of new ethnic identities, or through rapid acculturation leading to the adoption of a hybrid group identity, or through continuous socio-ethnic separateness resisting pressure to assimilate (the latter being typical of peripatetic, service-providing communities). Debatable however is whether or not the linguistic processes that lead to the emergence of MLs are unique to these situations, or commonplace. The

‘uniqueness’ approach is captured by the overwhelming tendency in the literature to assign notions and concepts to MLs that are not assigned to other linguistic contact phenomena: ‘intertwining’, ‘relexification’, ‘matrix language turnover’, ‘lexical re-orientation vs. selected replication’, and so on. Most authors agree that whatever the process, it draws on natural functional divisions in language and the speech production mechanism, in particular the division between the lexicon, and grammar. The need for tailored terminology expresses the need to capture a unique license which speakers adopt specifically in these social situations, to create and conventionalise a particular kind of balance, by source language, between these two principal components of the linguistic inventory.

This brings us to our two final issues. The first involves the debate on whether the processes that lead to the emergence of MLs are deliberate, in the sense that they reflect speakers’ conscious attempts to strategically manipulate language. The second is the extent to which the procedures of ML formation are regularised or pre-determined, in the sense that they will give rise to patterns of mixed structures that are at least to some degree predictable. Both questions are central to the ongoing Mixed Language debate.

4. Current issues in the Mixed Language debate

The questions addressed above might be summarised under three headings: The role of conversational mixing (codeswitching), the role of inherent divisions and layering within the language faculty, and the role of speakers’ intentions in the patterns of language mixing behaviour that give rise to MLs. These are the themes on which the present ML debate focuses.

Since MLs arise in situations of community bilingualism, it makes sense to search for their roots in patterns of language mixing at the level of conversation. Three pre-requisites are necessary, however, if we are to draw a connection between conversational codeswitching and the emergence of MLs. First, one needs to ascertain that codemixing was indeed an available strategy during the emergence stage. While this may be a common-sense assumption for some of the MLs, for others it is not at all obvious. Thus, many of the MLs which Smith (1995) has termed ‘symbiotic’ (since they are used alongside one of their source languages, usually alongside their grammaticiser language), are the outcome of just selective recruitment of lexicon from a language that was not actively accessible at the conversation level. This is true of Lekoudesch, for instance, where the Hebrew-derived lexicon was acquired from reading; and it tends to be true of most secret lexicons that have multiple

sources, such as Jenisch, and to some extent Ma'á or Abdal. It is also debatable whether MLs that may have emerged through imperfect learning or inter-language fossilisation, such as Javindo or Petjo, can be said to reflect earlier codeswitching patterns.

Next, it is necessary to demonstrate that the structural patterns of MLs are a reflection of a sedimentation of structural patterns of codeswitching. Myers-Scotton (1998, 2002, this volume) addresses the split between system and content morphemes, which is typical of both codeswitching at the utterance level, and most (though not all) MLs. But in this volume, Backus points out that insertional codeswitching is only part of the story in active bilingualism, and that MLs display no traces of alternational switching. While it makes sense that alternational switching cannot be conventionalised into a stable system, it is nonetheless noteworthy that the sedimentation of codeswitching patterns in an ML, if at all plausible, would entail giving up the option of alternational switching. The final point relates to the shift from insertional codeswitching as an occasional conversational option, ranging in frequency from low to moderate, to a conventionalised norm, that characterises each and every utterance. Muysken (1997) and Bakker (1997), as well as already Bakker and Mous (1994b: 5), argue that it is impossible to document the kind of continuum that Auer (1999) postulates, since codeswitching rarely reaches even approximately the extent of regularity in which the mixing of structures by etymology is conventionalised in MLs, nor does it normally affect the basic lexicon. In the centre of the debate on this issue is thus the problem of a continuum of quantity, and partly also of the quality, of insertions, and Backus (this volume) explores this argument further.

MLs have been acknowledged as a challenge to historical linguistics, but they are just as puzzling from the point of view of grammatical theory. Which kind of connections between grammatical categories are responsible for the (different patterns of) compartmentalisation of structures by source language in the various of MLs? Against the crude lexical:grammar split, which is not followed in all MLs, Matras (2000) had proposed that MLs arise due to different combinations of mixing processes, which are responsible for changes in different areas of grammar. In this volume, Matras follows up on that proposal and demonstrates the consistency of the language of the predication in all MLs, while Myers-Scotton argues that layering by source language is in part predictable through the natural layering of morphemes in the speech production and processing mechanism, as captured by the 4-M model. Both Stolz and Matras, as well as Bakker, discuss differences between heavy borrowing and MLs, Stolz focusing on the question of a possible continuum between the two.

Still disputed until fairly recently, there is now growing agreement that MLs are the outcome of deliberate processes of language manipulation. Statements to this effect were made already in the Bakker and Mous (1994) collection, and more recently by Matras (2000) and by Thomason (1999, 2001). In this volume, Thomason explains the contact mechanism ‘deliberate creation’, and Croft analyses the selection processes that speakers apply to elements of their bilingual repertoire in terms of positive and negative acts of identity. Golovko provides examples of speakers’ control over bilingual speech production which illustrate the ability to implement natural but conscious ‘language engineering’, and Mous surveys typical strategies of lexical manipulation and their manifestation in Ma’á in particular.

During the past decade, MLs have made their way from relative obscurity to inclusion in standard textbooks on contact linguistics. We hope that the present collection of contributions will further sharpen the focus on some theoretical aspects of the structure and emergence of MLs, and so allow MLs to further enhance our understanding of language and linguistic phenomena in general. We wish to conclude this introductory chapter by thanking the authors for their contributions to the debate, the series editors, Walter Bisang and Werner Winter, for offering this series as a platform for the debate, and the Arts and Humanities Research Board in England and Wales as well as the Spinoza fund awarded to Pieter Muysken for their support in bringing together the contributors and other colleagues for a face-to-face debate on Mixed Languages at the University of Manchester, in December 2000. We also thank the School of English and Linguistics at the University of Manchester and the Danish Institute for Advanced Studies in the Humanities (in particular Birgitte Possing), for their financial contributions towards the production of this book.

Abbreviations

CAUS	causative	OBV	obviative
DAT	dative	PL	plural
DM	discourse marker	POSS	possessive
F	feminine	REFL	reflexive
INSTR	instrumental	RESULT	resultative
NEG	negation	SG	singular
NOM	nominaliser	TOP	topic
NP	noun phrase	VP	verb phrase
OBJ	object		

References

- Arends, Jacques, Pieter Muysken, and Norval Smith (eds.)
 1995 *Pidgins and Creoles: An Introduction*. Amsterdam: John Benjamins.
- Auer, Peter
 1999 From code-switching via language-mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3: 309–332.
- Bakker, Peter
 1996 Language intertwining and convergence: Typological aspects of genesis of mixed languages. *Sprachtypologie und Universalienforschung* 49 (1), 9–20.
 1997 “A Language of Our Own”. *The Genesis of Michif – The Mixed Cree–French Language of the Canadian Métis*. New York: Oxford University Press.
- Bakker, Peter, and Mous, Maarten
 1994b Introduction. In Bakker and Mous (eds.), 1–11.
- Bakker, Peter, and Mous, Maarten (eds.)
 1994a *Mixed languages. 15 case studies in language intertwining*. Amsterdam: IFOTT.
- Bakker, Peter, and Muysken, Pieter
 1995 Mixed languages and language intertwining. In Arends, Muysken, and Smith (eds.), 41–52.
- Boretzky, Norbert, and Iglá, Birgit
 1994 Romani mixed dialects. In Bakker and Mous (eds.), 35–68.
- de Gruiter, Miel
 1994a Javindo, a contact language in pre-war Semarang. In Bakker and Mous (eds.), 151–159.
 1994b *Het Javindo. De Verboden Taal* [Javindo. The forbidden language]. The Hague: Moesson. (2nd, rev. edn.; 1st edn. 1990).
- Dreyfuss, G. R., and D. Oka
 1979 Chinese Indonesian: A new kind of language hybrid? *Papers in Pidgin and Creole Linguistics* 2. Canberra: Australian National University. (Pacific Linguistics, Series A – No. 57).
- Forman, M.
 1972 Zamboangueno texts with grammatical analysis. Ph.D. thesis, Cornell University.
- Frake, Charles O.
 1971 Lexical origins and semantic structures in Philippine Creole Spanish. In *Pidginization and Creolization of Language*, Dell Hymes (ed.), 223–242. Cambridge: Cambridge University Press.
- Golovko, Evgenij V.
 1994 Copper Island Aleut. In Bakker and Mous (eds.), 113–121.

- Grenoble, Lenore A., and Lindsay J. Whaley (eds.)
 1998 *Endangered Languages: Language Loss and Community Response*. Cambridge: Cambridge University Press.
- Kouwenberg, Silvia
 1994 *A Grammar of Berbice Dutch Creole*. Berlin: Mouton de Gruyter.
- Ladstätter, Otto, and Tietze, Andreas
 1994 *Die Abdal (Äynu) in Xinjiang*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Matras, Yaron
 1998 Para-Romani revisited. In *The Romani Element in Non-standard Speech*, Yaron Matras (ed.), 1–27. Wiesbaden: Harrassowitz.
 2000 Mixed Languages: A functional-communicative approach. *Bilingualism Language and Cognition* 3, 79–99.
 2002 *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Mous, Maarten
 1994 Ma'a or Mbugu. In Bakker and Mous (eds.), 175–200.
- Muysken Pieter
 1981 Halfway between Quechua and Spanish: The case for relexification. In *Historicity and Variation in Creole Studies*, A. Highfield and A. Valdman (eds.), 52–78. Ann Arbor: Karoma.
 1994 Callahuaya. In Bakker and Mous (eds.), 201–205.
 1997a Media Lengua. In Thomason (ed.), 365–426.
 1997b Callahuaya. In Thomason (ed.), 427–448.
- Myers-Scotton, Carol
 1998 A way to dusty death: The Matrix Language turnover hypothesis. In Grenoble and Whaley (eds.), 289–326.
 2002 *Contact linguistics*. Oxford: Oxford University Press.
- Sasse, Hans-Jürgen
 1992 Theory of language death. In *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*, Matthias Brenzinger (ed.), 7–30. Berlin: Mouton de Gruyter.
- Smith, Norval
 1995 An annotated list of creoles, pidgins, and mixed languages. In Arends, Muysken and Smith, 331–374.
- Thomason, Sarah G.
 1995 Language mixture: Ordinary processes, extraordinary results. In *Spanish in Four Continents. Studies in Language Contact and Bilingualism*, Carmen Silva-Corvalán (ed.), 15–33. Washington D.C.: Georgetown University Press.
 1997a Introduction. In Thomason (ed.), 1–7.
 1997b Ma'a (Mbugu). In Thomason (ed.), 469–487.
 1997c Mednyj Aleut. In Thomason (ed.), 449–468.

- Thomason, Sarah G.
 1997d A typology of contact languages. In *The Structure and Status of Pidgins and Creoles*, Arthur A. Spears and Donald Winford (eds.), 71–88. Amsterdam: John Benjamins.
 1999 Speakers' choices in language change. *Studies in the Linguistic Sciences* 29, 19–43.
 2001 *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. (ed.)
 1997 *Contact languages*. Amsterdam: John Benjamins.
- Thomason, Sarah G., and Terrence Kaufman
 1988 *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Vakhtin, Nikolai
 1998 Copper Island Aleut: A case of languages "resurrection". In Grenoble Whaley (eds), 317–327.
- van Rheedden, Hadewych
 1994 Petjo: The mixed language of the Indos in Batavia. In Bakker and Mous (eds.), 223–237.
- Wolff, John U.
 1982 Über die Gestaltung der Indonesischen Mundart der Peranakan-Chinesen in Ostjava. In *Gavac. Studies in Austronesian Languages and Cultures, dedicated to Hans Kähler*, R. Carle, M. Heinschke, P. W. Pink, Chr. Rost, and K. Stadlander (eds.), 451–461. Berlin: Dietrich Reimer Verlag.

Social factors and linguistic processes in the emergence of stable mixed languages

Sarah G. Thomason

1. Introduction

The title of this chapter assumes that “stable mixed language” is both a coherent category of languages and an actual phenomenon in the real world. Because both assumptions might well be challenged, I will try to make my own position as clear as possible (Section 1). In the main body of the chapter (Section 2), I will survey seven mechanisms of interference in an attempt to determine which ones are likely to have been relevant in the creation of these languages. The assumption underlying this survey is that, although mixed-language creation cannot be equated with language change, there is likely to be a great deal of overlap between the two kinds of process. In Section 3 I will draw some tentative conclusions, warning especially against simplistic one-size-fits-all approaches to the study of this unusual and fascinating class of languages.

To begin, then, mixed languages are in my opinion a coherent category. For me the term “mixed language” has a rather precise definition, one that does not coincide with that of some other authors (including several in this book). A principled definition of the notion is needed because, as has often been pointed out, all languages are mixed in a weak sense: there is no natural human language in which foreign material is wholly lacking.

Here’s my definition: a mixed language is a language whose grammatical and lexical subsystems cannot all be traced back primarily to a single source language. This definition is grounded in the historical linguist’s concept of genetic relationship, according to which each daughter language in a language family is a changed later form of its single parent language. It isn’t feasible to view a mixed language as a changed later form of any single parent language, because one or more of its component subsystems won’t fit into the standard genetic picture. It must be emphasized that there is no sharp boundary between “mixed language” and “unmixed language”. All languages have undergone at least some contact-induced changes, and in many languages, e.g. Asia Minor Greek (Thomason and Kaufman 1988: 215–222), there has been considerable restructuring through language contact – but without interrupting the main genetic lines, because systematic correspondences to sister

languages can still be found in all grammatical and lexical subsystems. (See below for further discussion, and see also Thomason 1997 for discussion of fuzzy boundaries in this domain, especially pp. 84–86, and for an argument that only a historical definition is adequate to identify mixed languages as a class; all the synchronic definitions that have been proposed to date fail to identify the category uniquely.)

It follows from this definition – and this is where I part company with some other authors interested in this topic (e.g. Bakker and Mous 1994; Bakker and Muysken 1995) – that the category of mixed languages includes pidgins and creoles as well as the set of languages designated by the rather clumsy term “bilingual mixed languages”. The fundamental division has to do with the processes by which mixed languages emerge: imperfect learning plays a significant role in the genesis of pidgins and creoles, but not in the genesis of bilingual mixed languages. This distinction correlates robustly with a sharp linguistic difference between the two types. In pidgins and creoles, the lexicon usually (but not always) comes mainly from one source language, but the grammar doesn’t; except for a limited number of specific features, the structural subsystems are most reasonably seen as a cross-language compromise among the languages whose speakers created the pidgin or creole. In bilingual mixed languages, by contrast, particular structural and lexical subsystems are adopted intact from each source language, sometimes (but not always) with a small amount of asymmetrical distortion or adaptation to the structure of the other language.

This neat typology of course hides a great deal of variation and fine grading from one type to another: the boundaries between one category of language and another are fuzzy, as is so often the case with historical linguistic categories – which are, by definition, subject to change. Scholars argue, for instance, about whether Afrikaans and Réunionese are creoles or direct descendants of Dutch and French, respectively (albeit very divergent ones that have undergone a great deal of shift-induced interference); the controversy focuses on the extent to which Dutch and French grammar are continued in the two later languages. Foreigner-talk Motu was apparently the foundation of the pidgin Hiri Motu (Dutton 1997), and there was almost certainly no sharp temporal or linguistic break between its status as a simplified kind of Motu and its later status as an independent pidgin. That is, it apparently developed by a gradual process, as has been proposed for various creoles by Chaudenson (e.g. 1992) and others. Berbice Creole Dutch, a rare two-language creole, has so much Eastern Ijò structure that it might be best analysed as positioned on the borderline between creole and bilingual mixed language. Here the crucial question concerns the issue of whether or not imperfect learning contributed sig-

nificantly to the genesis process. But although there are fuzzy boundaries and therefore borderline cases, there are also many clear cases – cases involving abrupt creation and therefore sharp breaks in transmission, cases of clear mutual unintelligibility between mixed language and lexifier language (where there is a simple lexicon vs. structure division), and so forth.

That, in brief, is my position on mixed languages as a general category. In the rest of this chapter, I'll focus on the subcategory of bilingual mixed languages. Here is where the question arises as to whether stable mixtures of this type actually exist in the real world. In one sense, of course, no language is stable: all living human languages are constantly changing, so that the only truly stable language is a dead one. And stability in the sense of continuing existence is an accident of history: languages do not live or die according to their merits or their linguistic components, but according to the fates and choices of their speakers. These kinds of linguistic and social instability are shared by all languages, however, and so are of no special interest when the topic is mixed languages.

There is no doubt that bi- or multilingual speakers, under many circumstances, often combine parts of two (or, sometimes, more than two) languages in speech, or that some of these combinations achieve a social status that deserves the label “stable”. In practice, however, there is frequently considerable controversy over whether such combinations are languages or merely ephemeral combinations. If they are languages, they must be learned as such – learned as entire interconnected complexes of lexicon and structure, though not necessarily by children as a first language. That is, to count as stable mixed languages, they must not be produced as independent ad-hoc combinations by each speaker in each mixed utterance. (I will resist the temptation to go into the related question that some scholars have raised – namely, whether there is any such thing as “a language” – except to observe that stable mixed languages have as much claim to the label as other languages.) Few bilingual mixed languages enjoy a completely undisputed status as fully crystallized languages; disagreements are the norm, not the exception. Media Lengua, for instance, is described by Muysken (e.g. 1997) as a stable mixed language, but at least a few other scholars are skeptical about its stability (Rodolfo Cerrón-Palomino, p.c. 1997), viewing it instead as an ephemeral and unstable phenomenon. Older descriptions of Ma'á present the language as a stable mixture (see e.g. Whiteley 1960; Goodman 1971; Tucker and Bryan 1974) and its speakers as a separate people (e.g. Kimambo 1969; Feierman 1974); but Mous (e.g. 1994 and 2000 ms.), working from later field data, finds Ma'á to be a mere sociolect of the Bantu language Pare, and its speakers to live with Shambaa speakers as their near neighbours, not as a completely separate community.

It's probably fair to say that the only uncontroversially stable bilingual mixed languages are those that are now spoken outside the bilingual context in which they arose. So, for instance, most speakers of Michif on the Turtle Mountain Reservation of North Dakota are fluent neither in Cree nor in French; and Mednyj Aleut continued in use during the period when the Russian presence was largely withdrawn from the Mednyj (Copper) Island. Both of these enjoy a reputation as stable mixed languages, despite the existence, especially in the case of Michif, of (partly) similar but less stable and less mixed speech varieties in other Métis communities.

The status of Kormakiti Arabic, which is spoken by Arabic-Greek bilinguals on Cyprus, is a typical illustration of the stable-vs.-unstable controversy. Newton describes it as a stable mixture (1964), and Jastrow (1977) seems to agree (though his position on the question of stability is not entirely clear). In sharp contrast, Borg concludes that, "far from being a 'mixed language' (pace Newton 1964), KA has . . . largely remained a separate if receding linguistic system" (1985: 153). Borg argues that Kormakiti Arabic was never a mixed language, but instead an elaborate codeswitching phenomenon in which Greek elements appeared in the speakers' Arabic in quantities dependent on social context – more Greek in more formal contexts, thanks to the low prestige of Arabic on Cyprus. But the people Borg interviewed were not the same population that Newton had worked with: by the time Borg began his research on Cypriot Arabic, political events had made it impossible for outsiders to visit the village of Kormakiti, and in addition many or most of the speakers had scattered to other regions. Borg's evidence is therefore not fully convincing, at least not to me; it seems quite possible that Newton's original assessment of the linguistic situation was correct, for the speakers he worked with. If so, then – as was quite likely the case with Ma'á, where the picture as of the 1960s apparently differed significantly from the situation fifteen or twenty years later – Kormakiti Arabic may well have undergone a rapid decline in the years between Newton's visit and Borg's. The language Borg investigated resembles in some ways the language Mous reports, greatly reduced in its grammatical, lexical, and functional range, and in fact replaced by the language that had become dominant for the speakers of the dying language, all of whom had long been bilingual.

Another important question that arises in this context is whether there is anything linguistically special about stable mixed languages, as opposed to unstable mixtures. For instance, Michif is typically considered to be a prime example of a stable mixed language (see e.g. Bakker and Papen 1997), but essentially the same linguistic mixture found in Michif is apparently found in at least one unstable Algonquian–French contact medium in Canada (Drapeau

1991). Moreover, the noun-phrase-verb-phrase dichotomy of Michif resembles some non-fossilized codeswitching patterns elsewhere. An example is French-Arabic codeswitching in Morocco (Heath 1989: 38), in which French noun phrases are much more common than complete French verb phrases as code-switches. In this case, achieving stability apparently did not entail any special linguistic processes; and for other stable mixed languages too it is easy to find linguistic analogues in less stable contact phenomena (Thomason 1995).

If the stability of a mixed language is not established by any unique linguistic properties, then the difference between stability and instability in cases of bilingual language mixture must be a function of social differences. My main argument in this chapter, in fact, is that social factors determine stability. To explore the question of how stable mixed languages arise, I will survey the seven mechanisms of interference that I've identified (Thomason 2001: ch. 6), attempting to determine which of these are likely to have been operative in the development of stable mixed languages. This survey in turn leads to conclusions about the nature of bilingual mixed languages. Historically, some arose relatively abruptly, as symbols of new ethnic groups, while others arose, probably over a longer period of time, as minority ethnic groups clung to their old cultural identity, resisting total linguistic assimilation to a dominant group. But it is surely premature to draw firm conclusions about what linguistic and social processes can and can't produce bilingual mixed languages: the number of well-understood languages of both types is so very small that we still don't have much to go on.

To close this section, another type of challenge to the existence of mixed languages should be mentioned briefly. Greenberg claimed that

It is indeed hard to imagine how a truly mixed language in . . . the usual sense, could arise by a natural process. Suppose someone had a dictionary and grammar of two quite distinct languages. He or she could then take alternate words and grammatical morphemes first from one and then the other. This would truly be a mixed language but, of course, not arising by any natural process. (Greenberg 1999: 632)

Greenberg's odd definition of "a truly mixed language" is hardly "the usual sense", as anyone familiar with the literature on bilingual mixed languages would know: Greenberg is, to the best of my knowledge, the only author who has ever suggested that the only kind of mixture that would count as a bilingual mixed language would have to be formed by taking alternate words and morphemes first from one language and then from the other. His

article contains no arguments whatsoever in support of his idiosyncratic characterization. What we find instead, in the real world, is a variety of mixtures that divide according to grammatical subsystems – often lexicon vs. grammar, but in other ways too, for instance noun phrases vs. other subsystems (Michif) and finite verb inflection vs. other subsystems (Mednyj Aleut). In a limited sense, however, Greenberg's genesis scenario is perhaps less far-fetched than he imagined, because (as I'll argue below) deliberate creation appears to have been a major contributor to the development of a number of mixed languages. Still, because Greenberg's scenario would require sustained intensive creative effort by highly literate creators, it isn't surprising that it apparently isn't instantiated anywhere. A question remains as to whether he is right in claiming that it wouldn't be a "natural process": he offers no definition of "natural process". Some of us would argue that anything speakers do to their language, including deliberate manipulation, must be a natural process; and, as is increasingly clear from the growing number of instances of linguistic manipulation by speech communities, the borderline between unconscious change (which historical linguists have always considered the norm) and conscious, deliberate change is highly permeable.

2. What mechanisms of interference contribute to the development of bilingual mixed languages?

This question is the main topic of the present chapter. I do not suggest that ordinary contact-induced language change can be equated with the creation of a bilingual mixed language; it can't be. But by the same token, there is no reason to assume that speakers' behaviour under ordinary contact conditions will be wholly different from their behaviour in the extraordinary kinds of contact conditions that lead to the emergence of a bilingual mixed language. It seems reasonable, then, to look for congruences between processes of contact-induced change and processes of mixed-language genesis. As far as the strictly linguistic processes are concerned, there is certainly no sharp divergence between ordinary contact-induced changes and the combinations of features found in bilingual mixed languages: the difference is in degree, not in kind. The ordinary-change mechanisms are therefore a good starting point for a discussion of processes of mixed-language genesis. (See Thomason 1995 for a more detailed discussion of links between processes of mixed-language genesis and processes of contact-induced language change.)

Another point that must be emphasized is that there are no direct observations of any linguistic processes that have led to the emergence of stable

mixed languages, so that we are limited to indirect evidence in theorizing about modes of origin. Indirect evidence, as I'll try to show in this section, does permit some cautious conclusions about mechanisms that are likely to have been operative in the processes through which various mixed languages have arisen.

The seven mechanisms are codeswitching (in a broad sense, including both intersentential and intrasentential switching), code alternation, passive familiarity, "negotiation", second-language acquisition strategies, bilingual first-language acquisition, and change by deliberate decision. I'll survey each of them in turn.

Codeswitching is very likely to have been an important mechanism, and perhaps the most important mechanism, in the genesis of Michif. As noted above, the particular mixture in Michif – noun phrases vs. the rest of the language – matches at least some other codeswitching patterns quite closely, including Moroccan Arabic–French switching and also, closer to the context in which Michif arose, patterns involving Algonquian languages: Plains Cree–English (Bakker 1997: 181–182) and Montagnais–French (Drapeau 1991, cited in Bakker 1997: 184–186). Bakker in fact observes that “. . . all cases of code mixing that involve Algonquian languages are of the insertional type . . . overwhelmingly constituent insertion: whole noun phrases . . . and prepositional phrases from English and French are inserted into the Algonquian framework” (1997: 180). This doesn't of course mean that codeswitching was the only mechanism involved, but it seems to have provided the foundation on which the crystallized mixed language was built. I'll argue below that the seventh mechanism, change by deliberate decision, is also likely to have been an important factor in its genesis.

This proposal about Michif goes against the view of Peter Bakker, who has written most extensively about this language (in particular in Bakker 1997, but elsewhere as well, e.g. in Bakker and Papen 1997). Bakker offers four main arguments against the hypothesis that codeswitching – or, in his terminology, code-mixing – played a major role in the emergence of Michif. First, he asks whether nineteenth-century Cree–French bilinguals could have been code-mixing, and concludes that “[t]his is unlikely . . . since nowhere in the world are there reported cases of code mixing [sic] in which only certain categories (e.g. only nouns) are taken from the other language” (1997: 11). Second, he argues that, because present-day Michif speakers “rarely know Cree and usually do not know French either”, “[i]n a contemporary analysis, Michif cannot be an instance of code mixing” (1997: 22). Third, in spite of the apparent parallelism between Michif and Montagnais–French codeswitching or borrowing, “[w]hereas all nominal items are French in Michif, in Montagnais

they would be limited to . . . ‘cultural borrowings’” (1997: 186). And fourth, “Michif is not a case of fossilized code mixing” because “French elements have properties of the Cree items that were replaced by the French nouns” (1997: 190).

Some of Bakker’s objections to a codeswitching as a mechanism in the genesis of Michif would be relevant only to a claim that Michif is purely and simply fossilized codeswitching, with no contributions from other mechanisms. I’d agree with him that this simplistic scenario is implausible. But it seems to me that his absolute either/or approach obscures important parallels between the structure of Michif and a variety of codeswitching phenomena. I’ll discuss each of his four arguments in turn. He first argues that the bilingual creators of Michif were probably not codeswitching because the eventual mixture in the fully-crystallized contact language is not identical to what is found in any codeswitching context. But in fact the mixture does closely resemble the noun phrase insertional codeswitching found in a variety of other contexts (including the ones I’ve cited above); and in any case, there is no reason to expect the eventual structure of the mixed language to be identical to any fluid codeswitching context. His second argument – that current speakers of Michif can’t be codeswitching because they don’t know both languages – is not relevant to a discussion of how Michif arose in the first place. Everyone agrees that Michif was created by people who spoke (at least) French and Cree fluently. Differences between the structure of Michif and either codeswitching or borrowing in Montagnais would be relevant only if one assumed that the codeswitching that (by hypothesis) helped establish the structure of Michif had to be completely identical to current codeswitching in Algonquian languages, and also that codeswitching had to have been the only mechanism in the genesis of Michif. Similarly, Bakker’s fourth argument would be telling only in the context of two controversial beliefs: that codeswitching can only have been relevant if it was the sole mechanism employed by the language’s creators, and that codeswitching and borrowing and other contact phenomena can safely be assumed to be rigidly compartmentalized, both semantically and formally.

Media Lengua may also have arisen at least in part via codeswitching. Muysken notes that bilingual Quechua–Spanish songs that are popular in Peru and Bolivia involve frequent codeswitching and borrowing as well as the process he distinguishes as relexification (1997: 418), and he concludes his study of Media Lengua by pointing to the need to understand the interaction between the processes through which Media Lengua arose with “other contact processes such as codeswitching . . .” (1997: 420). For him, apparently, the question of whether codeswitching was one of the genesis mechanisms in-