INTRODUCTION

Peter Bakker & Maarten Mous

This is, as far as we know, the first book devoted to mixed languages ever published. It is the result of a workshop on mixed languages held in Leiden, the Netherlands, on the 24th of May, 1993 and organised by Peter Bakker (University of Amsterdam) and Maarten Mous (Leiden University). We were both working on individual mixed languages and were interested in the general problem of mixed languages. We invited linguists working on other mixed languages to come to Leiden in order to compare the various characteristics of these languages and to discuss the related theoretical issues. The agenda for the discussion was determined by the participants, as they formulated the questions that they wanted to be on the agenda beforehand. We classified and distributed these questions so that the participants could study them and prepare answers to be given at the workshop. These questions were the following:

Structure: the nature of the mixture.
For each language, what comes from which language?
(general)
- grammatical system versus lexicon?
- are functional or semantic aspects relevant as the source of lexical or grammatical items?
(morphology)
Is there any difference between:
- prefixes versus suffixes?
- derivation versus inflection?
- productive/ half-productive /unproductive morphology?
- accentuated versus non-accentuated morphemes?
(phonology)
- which language is the source of the phonological system?
(syntax)
- which language is the source of the word order rules?
- which language is the source of free grammatical morphemes?

Structure: questions of genetic relationship and typological distance of source languages.
- How is typological distance between the source languages overcome in the mixed language?
- When two tongues which are very similar and have similar surface structures come into contact, it seems that grammatical morphemes travel easily (e.g. between dialects). When two less similar tongues (two languages from the same family or two languages from different families) come into contact, one finds that grammatical categories are transferred less easily, e.g. tense-aspect categories. Is that the experience of the other researchers?

Process(es) of genesis.
- Are there any indications as to the direction of the foreign influence, lexical or grammatical borrowing?
- How common is massive borrowing? ('I have been investigating languages with 10,000 to 20,000 speakers. All have borrowed extensively from their neighbours. For each, the available lexicon is between 1000 and 2000 lexemes and/or stems. Between 50 % and 75 % of each lexicon is transferred = borrowed. Is that common?')
- Is the language spoken by a group with a separate identity?

Community: The mixed language in a wider geographical and social context.
- Is the language intelligible for speakers of both source languages?
- How intelligible or recognizable is the language for speakers of either of its source languages?
- Is it an in-group language?
- Is it different from code-mixing between the same languages?
- How is the attitude of the speakers themselves towards the mixed language, and how is the attitude of the speakers of either of its source languages towards the mixed language?

Language acquisition aspects.
- Were children, adolescents or adults responsible for the 'creation' of the mixed language?
- Did the language come into being suddenly or slowly?
- Are the language components simplified as to their source languages?
- Are first language acquisition and/or second language acquisition responsible or relevant for the genesis of the mixed language?

Genesis: reconstruction of the social context of genesis.
- Does a separate identity of the group of speakers play a role in the genesis of the mixed languages?
- Did the first generation of speakers have fathers speaking a language different from the mothers?
- Are language death processes responsible for the mixed nature of the language?
- Did code-mixing play a role in its genesis?
- Was it at first a contact language between speakers of the source languages?

Mixed languages and language contact models.
Open discussion.

Mixed languages as a type: similarities and dissimilarities.
- How can we identify a mixed language?
- Are there any typological similarities within the group of mixed languages?
- Thomason & Kaufman (1988) propose four types of contact-induced language change: shift, borrowing, pidginization and abrupt creolization. Is there a fifth model called 'language intertwining' which is responsible for the genesis of mixed languages, in which a grammatical system and a lexicon are combined into a new language?
- According to T & K these models ought to be recognizable retroactively by the traces they leave. 'I have been trying to find clear cases of the types in East Africa and have come to the reluctant conclusion that the kind of contact that leads to borrowing is much more common than the other types. Is that the universal experience?'
- What are the typological similarities / differences between mixed languages and other types of contact-induced languages?
- Is it useful to talk about non-genetic or multi-genetic developments in historical linguistics?

Evaluation: How, why and when do mixed languages come into being?
Open discussion.

The participants prepared a one-page overview of their language, which was distributed to all participants. Each cluster of questions was discussed for each language by the linguist who studied this language. Some answers were as short as 'I don't know' or 'Not relevant', but for most of the time turn-taking had to be regulated by our chairpersons Willem Adelaar (Leiden University) and Jacques Arends (University of Amsterdam) and the success of the workshop can for a large part be attributed to their excellent fulfilment of this difficult task. The participants were very satisfied with the discussions and the structure of the workshop. It provided new avenues for their ongoing research. In a way it is unfortunate that we cannot revive the discussion here.

Instead, this collection contains contributions by all the participants of the workshop. Moreover, we were able to obtain three additional articles. Antony Grant and Maarten Kossmann, both present at the workshop, found inspiration to each write an article on other mixed languages, Shelta and Amarna Akkadian respectively. The paper on mixed varieties of the Romani (Gypsy) languages was requested by the editors in order to fill an obvious gap in the set of languages that were represented at the workshop. Boretzky and Igla's paper was originally written for another volume, but remained unpublished. We are pleased that the authors agreed to publish their paper in this collection. It fits very well into the fact-finding character of this collection and it also contains interesting insights which may account for the emergence of these mixed languages.

All in all, the materials presented in this book have a very broad geographical and temporal span. The oldest language dealt with is Amarna Akkadian recorded on clay tablets from around 1400 BC; the newest language is Media Lengua, a mixture of Quechua and Spanish which must have come about only in this century. The geographical spread is just as vast. Two languages are from South America (Callahuaya and Media Lengua), one from the Caribbean (Island Carib), one from North America (Michif). Six languages are from Asia: Petjo and Javindo from Indonesia and Copper Island Aleut from the Bering Strait, two Romani mixed varieties from the Near East, Qirishmal and Armenian Romani, as well as Amarna-Akkadian. Three languages are from Africa (Ilwana, KiMwani and Ma'a/Mbugu). From Europe we have Maltese, Shelta (the language of the Irish Tinkers), Stedsk or Town Frisian from the Frisian part of the Netherlands, Angloromani in the British Isles, and several other Romani mixed varieties such as Norwegian Romani, Dortika in Greece, Basque Romani in the Basque Country, Šatrovački in the Balkans and Caló in Spain. The fact that some geographical areas are missing (such as Australia and Papua New Guinea) is due to chance: there are also reports on mixed languages from these regions.

Some of the languages discussed in this volume and possibly also other mixed languages, will also be dealt with in a forthcoming book on non-European contact languages edited by Sarah Grey Thomason.

1. The definition of 'mixed language'.

The definition of a mixed language is not without problems. As is the case with the definition of creole languages, the definition of a mixed language is a theoretical issue. We propose the term 'language intertwining' for the process forming mixed languages showing a combination of the grammatical system (phonology, morphology, syntax) of one language with the lexicon of another language.

There is a long tradition of discussion on the theme of mixed languages, going back to the 19th century (some of it summarized in Thomason & Kaufman 1988). The result of the discussion was that the idea that mixed languages do not exist gained general acceptance. This idea was partly due to the fact that languages such as English and creole languages were discussed, neither of which would qualify as mixed languages in our view. Nevertheless, the idea that mixed languages do not exist has to be reconsidered, and we think that the papers in this volume show convincingly that mixed languages do exist. Nevertheless, it is not easy to formulate a definition.

Attempts at a definition of a mixed language can be found in the contributions by Bakker and Golovko. A very rough approximation is that a mixed language has its lexicon and grammar from different sources. On the basis of the lexicon one would classify such languages as belonging to one language family and on the basis of the morphology, syntax and general grammatical characteristics one would classify them as belonging to another language family.

As languages are classified both on the basis of the lexicon (especially the core vocabulary) and on grammatical features, notably morphological similarities, most of the languages discussed in this volume cannot be classified as belonging solely to family A or solely to family B, and thus they are mixed languages. They have genetic connections with two language families or branches.

One obvious problem is that such is a matter of degree and it is not *a priori* clear what qualifies as a mixed language and what as a case of extreme borrowing. Ilwana (Nurse) and especially Maltese (Drewes) may be instances of the latter.

One point where one could distinguish between extreme borrowing and mixed languages are the proportion of lexical items that are of 'foreign' origin (i.e. from the perspective of the language of the grammatical system). A second point would be whether these 'borrowed' items are in the core vocabulary or are peripheral to this. As for the proportion, one can see that extreme borrowing never exceeds roughly 45 % of the lexicon, whereas in some of the mixed languages discussed the proportion of 'foreign' lexical elements is closer to or over 90 %, and this figure is the same whether one counts types or tokens. There do not seem to be languages with a proportion of borrowed items between 45 % and 90 %, so that there is no continuum between languages with heavy borrowing and mixed languages. This suggests that the mixed languages are different from extreme cases of lexical borrowing (whereby Maltese would be a case of the latter). The other point is that in the mixed languages most of the core vocabulary tends to be foreign,

whereas in cases of extreme borrowing the foreign elements affect the core vocabulary only to a limited degree.

Another pitfall for definitions of mixed languages is the fact that the linguistic history needs to be known, at least in part, before one can decide whether a given language qualifies as a mixed language or not. That is to say, the two sources for grammar and lexicon need to be recognised. This may be problematic not only because of the lack of knowledge of the linguistic history, but also because one of the source languages may be extinct. Only in exceptional cases is it possible to recognize a language as being mixed without knowledge of the source languages. Michif, with its two phonological systems, would probably be recognized as mixed even by a linguist unfamiliar with the source languages.

In addition, the presence of sufficient affix morphology seems to be a prerequisite for recognising that the source language of the grammar is different from the lexical source. But are the processes that lead to a mixed language really limited to morphologically rich languages? It does not seem to be so, witness the existence of Petjo having the morphologically poor language Malay as its grammar language.

Another point of concern is that the source languages are preferably far apart, typologically or genetically. This prerequisite has a practical side, that is: if far apart, the different source languages are more easily recognised. On the other hand, certain historical processes may occur precisely when speakers are able to equate comparable morphemes, i.e. when the source languages are similar. KiMwani (Schadeberg), Stedsk (Van Bree), and to a lesser extent Amarna Akkadian (Kossmann) are examples for which the source languages are related. In the case of KiMwani two Bantu languages, in the case of Stedsk two West Germanic languages and in the case of Amarna Akkadian two Semitic languages. In all the other cases, the mixture involves two languages which are not genetically related at all, or at most remotely related.

And then there is the question of the number of languages involved. In the first approximation of the definition mention was made of two source languages, a grammar language and a lexicon language. But some of the languages in this collection have more than one lexical source, in a substantial way, see for example Ma'a (Mous), Maltese (Drewes), Shelta (Grant).

A term 'mixed language' implies that there is such a category. The question is: Do the languages represented in this book constitute one group with common characteristics or do they form a collection of oddities among the languages of the world? We see several observations that are made again and again for individual cases. Still one might want to distinguish subcategories. Bakker distinguishes two types of mixed languages on social grounds, roughly

mixed languages that developed out of mixed marriages and secret languages of nomadic people (Bakker 1992, chapter 7). Thomason makes a division between mixed languages that have developed gradually as opposed to those that developed abruptly, the latter ones serving a newly formed social group (Thomason to appear). Smith (1994) distinguishes between plain mixed languages and symbiotic mixed languages. The former are generally the only language spoken in a community, whereas the latter coexist with dominant unmixed languages. The latter may also have a more limited lexicon. Other variables along which these languages can be divided are equally possible.

This collection may serve as a test ground for attempts at a definition of mixed languages, for a taxonomy of the same and for deliberation on the desirability of the concept.

2. Scarcity of data.

Data on mixed languages are scarce. We hope that this collection rectifies this situation to some extent. Apart from the fact that little study has been done on the vast majority of languages in general, the low status of most mixed languages - they are not always regarded as full fledged languages -, the fact that the mixed languages are often spoken by small or outcast groups (e.g. nomadic groups such as Gypsies) further contributes to this scarcity of data. Not to mention the fact that some of these languages are considered as secret languages by their speakers, such as Dortika and Angloromani (Boretzky & Igla), Callahuaya (Muysken) and Shelta (Grant).

3. Comparative questions.

We will mention here only some of the structural factors that are shared by several of the languages. Where we mention examples for certain phenomena, these lists need not be exhaustive. Our statements may be overgeneralised and simplified, but the purpose is to direct the reader to the articles.

As far as the grammar language is concerned, most languages in our collection have one source for the grammar. A limited number show a mixture in the grammar, a combination of the lexicon language and the other source language, Stedsk (Van Bree) for example. And there are the spectacular cases in which the sources of the grammar are divided structurally, such as Michif (Bakker) that on first inspection has the nominal grammar from French and the verbal grammar from Cree, and Copper Island Aleut (Golovko) which has Russian pronouns and verb suffixes in the otherwise Aleut verb phrase.

In a number of cases, the community not only speaks the mixed language

but is also fluent in the language that is the grammar source, such as Javindo (De Gruiter), Petjo (Van Rheeden), Island Carib (Hoff), and Ma'a (Mous) and the mixed Romani dialects, but this is not the situation in Michif (Bakker), Ilwana (Nurse) and KiMwani (Schadeberg). In other cases speakers of the mixed language can also speak the language of the lexicon. This is true for at least some Media Lengua speakers (Muysken) and quite a few Michif speakers (Bakker).

The mixed languages that arose out of mixed marriages take their grammar from the language of the mother. These are the following: Michif (Bakker), Javindo (De Gruiter), Petjo (Van Rheeden), Island Carib (Hoff) and possibly Copper Island Aleut (Golovko).

For Romani (Boretzky & Igla), Stedsk (Van Bree) and Callahuaya (Muysken), the grammar source is the language that the speech community seems to have shifted to. In Media Lengua (Muysken), however, Quechua is the source for the grammar while the population is shifting to Spanish.

Some mixed languages show simplifications in their grammar when compared to the source, see for example the lack of inflection on the verb in the Greek-Romani or Dortika (Boretzky & Igla), and Copper Island Aleut (Golovko).

Closed sets of function words are often exceptional in that they are not (all) from the grammar source. Either they are of mixed origin or they are from the lexifier language. Boretzky & Igla mention 'small word classes' quite often as being of Romani origin (specifically pronouns are derived from Romani inflected forms) with additions from the grammar language. In Copper Island Aleut (Golovko) too, the closed sets in part are from the lexical source. In Ma'a (Mous) demonstratives, possessives and personal pronouns have the categorisation but not the actual content form from the grammar source. Petjo has 'free grammatical morphemes' from Dutch in an otherwise Malay grammar. Michif shows variation in the source language of some free grammatical morphemes, which differs somewhat from community to community.

Interestingly, in Stedsk (Van Bree) and Shelta (Grant) stressed and non-stressed morphemes come from different source languages.

Some mixed languages are described as a lexical register. Island Carib (Hoff) is a clear case: the mixed register was used when addressing males and the non-mixed register for addressing females, since the males and females had to be spoken to in their own language. Ma'a (Mous) too can be described as a lexical register of a non-mixed Bantu language. The characteristics of registers, such as the presence of parallel word forms for one and the same meaning, can also be discerned in other languages. The article on Maltese (Drewes) shows in detail all sorts of calques, and similar things are reported for Petjo (Van Rheeden).

Most authors who compare the mixed language with code switching between the grammar language and the lexifier language claim that there is a clear structural difference between the two phenomena: Golovko for Copper Island Aleut, Grant for Shelta, Muysken for Callahuaya.

Some languages are not a first language but are acquired at a later age: Angloromani (and probably some of the other Romani mixed languages as well) (Boretzky & Igla), Javindo (De Gruiter), Petjo (Van Rheeden), Amarna-Akkadian (Kossmann). In the article on Petjo (Van Rheeden), a link is made to second language acquisition.

A number of mixed languages are threatened by extinction, or are already extinct. This is particularly the case for the mixed languages of which the communities are also fluent in the grammar language. Several Romani varieties (Boretzky & Igla) are threatened by extinction, Stedsk (Van Bree) is threatened by Dutch. Other mixed languages seem to be quite stable, for example KiMwani (Schadeberg) and Ma'a (Mous). Some languages are limited in function, e.g. Callahuaya (Muysken) which is the ritual language of the healers.

4. The sociohistory.

Some recurrent themes about the sociohistory of the languages discussed in this collection are presented in the following.

The mixed language is an in-group language, e.g. Copper Island Aleut (Golovko), Javindo (De Gruiter), Petjo (Van Rheeden), Shelta (Grant), and Callahuaya (Muysken).

Speakers have a positive attitude to the mixed language in the case of Copper Island Aleut (Golovko), Michif (Bakker), Maltese (Drewes), Amarna-Akkadian (Kossmann), Media-Lengua (Muysken). In some situations, there is a negative attitude to the dominant population or a strong urge to be different: Javindo (De Gruiter), Petjo (Van Rheeden), Ma'a (Mous), Shelta (Grant).

The creators of mixed languages can sometimes be characterised as social risers, as is the case for KiMwani (Schadeberg), Copper Island Aleut (Golovko), or Stedsk (Van Bree); or as male adolescents, Media Lengua (Muysken), Petjo (Van Rheeden), Javindo (De Gruiter); or as educated: the scribes that used Amarna Akkadian (Kossmann), and the monks who were possibly influential in the making of Shelta (Grant).

Language shift or language death is sometimes but not always mentioned as a factor in the origin of the mixed language, but not always.

Mixed languages do not need to be contact languages, in the sense that they were used in contacts between different language groups. More often

than not mixed languages did not come into being because of communication needs (with possible exceptions like Stedsk), as its speakers also had or have other languages in common.

Many contributions mention a deliberate creative element. Boretzky & Igla mention a wilful avoidance of admixture to the lexicon with words from the grammar language for the Romani mixed varieties, and Shelta (Grant) shows deliberate creation of lexemes through various transformations. The same is observed for Ma'a (Mous).

5. Conclusion.

The case studies in this book form a valuable challenge and contribution to historical linguistics. The changes are spectacular in outcome and - when recoverable - in speed. The correlation of the mixture and the social context is crucial in the understanding of the processes involved. We hope that the contributions to this book will help to make the study of mixed languages and their genesis a respectable field within historical linguistics and sociolinguistics. They show the manifest necessity in the study of these languages that historical and social factors should be linked with structural and typological properties of the languages involved. Only in this way can one try to unravel the genesis of these mixed languages. The social factors, and not linguistic ones, are responsible for the emergence of mixed languages. We further realize that languages like these (especially in their dichotomy of grammar and lexicon) challenge various assumptions made in psycholinguistic and formal linguistic theories.

The comparative study of mixed languages is only just beginning. This book provides materials and analyses on individual mixed languages rather than a unified theory. Only with the availability of these materials can a serious comparative study begin.

References

BAKKER, Peter
1992 *'A Language of our Own': The Genesis of Michif, the Mixed Cree-French Language of the Canadian Métis.* Ph.D. thesis, University of Amsterdam. (to be published by Oxford University Press, 1995).

SMITH, Norval
1994 'An annotated list of Creoles, Pidgins, and Mixed languages'. In J. Arends, P. Muysken & N. Smith (eds.) *Creoles and Pidgins. An Introduction.* Amsterdam: John Benjamins.
THOMASON, Sarah Grey
t.a. Language mixture: Ordinary processes, extraordinary results. Ms., University of Pittsburgh.
THOMASON, Sarah Grey (ed.)
f.c. *Non-European Contact Languages.* (working title).
THOMASON, Sarah Grey & Terrence KAUFMAN
1988 *Language Contact, Creolization and Genetic Linguistics.* Berkeley: University of California Press.

Acknowledgements

We want to thank Eithne Carlin and Inge Genee for their editorial advice on some of the papers. Neither bears any responsibility. We also want to thank IFOTT, Computer Centrum Letteren and the Institute for General Linguistics for the use of their computers, conversion software, scanners and printing facilities.

The workshop was supported financially by Stichting Taalwetenschap and IFOTT, the Institute for Functional Research of Language and Speech. To both we express our gratitude.