

BILINGUAL SPEECH
A TYPOLOGY OF CODE-MIXING

PIETER MUYSKEN



PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

© Pieter Muysken 2000

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Times 9.5/13pt [CE]

A catalogue record for this book is available from the British Library

ISBN 0521 77168 4 hardback

CONTENTS

<i>List of figures</i>	<i>page</i> vi
<i>List of tables</i>	vii
<i>Preface</i>	xi
<i>List of abbreviations</i>	xiv
1 The study of code-mixing	1
2 Differences and similarities between languages	35
3 Insertion	60
4 Alternation	96
5 Congruent lexicalization	122
6 Function words	154
7 Bilingual verbs	184
8 Variation in mixing patterns	221
9 Code-mixing, bilingual speech and language change	250
<i>References</i>	279
<i>Author index</i>	298
<i>Subject and language index</i>	304

FIGURES

1.1	Schematic representation of the three main styles of code-mixing and transitions between them	<i>page 9</i>
1.2	Schematic representation of two ways of conceptualizing the relation between various mixing patterns	31
1.3	The present classification of language interaction phenomena as compared with that of Myers-Scotton and Poplack	32
8.1	Localization of a number of contact settings involving Dutch in the triangle alternation, insertion, congruent lexicalization	245
8.2	Localization of a number of contact settings in the triangle alternation, insertion, congruent lexicalization	246
8.3	Factors governing the choice of a particular code-mixing strategy in different communities	248
9.1	Typology of language contact phenomena	265

TABLES

1.1	Classification of code-mixing theories in terms of the notions of asymmetry, word order, categorical equivalence, peripherality, and functional elements	<i>page</i> 28
3.1	Features of code-mixing and borrowing according to Sankoff and Poplack (1984)	73
3.2	Language choice for animal names, differentiated for grammatical context and morphological shape (based on Muysken, Kook, and Vedder 1996)	91
4.1	Number, percentage of total number of borrowings, and type–token ratio (TTR) of French words in Brussels Dutch (Treffers-Daller 1994)	99
4.2	The use of French discourse markers in Shaba Swahili/French bilingual speech	112
5.1	Examples of two paradigms in sociolinguistic research: the variation paradigm and the code-mixing paradigm	127
5.2	Multi-word switches in the Ottersum corpus (based on Giesbers' tables 4.3.2 and 4.3.5)	130
5.3	Single-word switches in the Ottersum corpus (based on Giesbers' tables 4.3.2 and 4.3.5)	131
5.4	Single Dutch elements in the Sranan corpus (based on Bolle's tables 5.2 and 5.11)	139
6.1	The classification of Dutch elements in terms of a number of the criteria listed	159
6.2	Myers-Scotton's system for distinguishing between system (s) and content (c) morphemes (1993b: 101)	162
6.3	Functional elements particularly affected in different linguistic domains	171
6.4	English verbs in a Spanish grammatical context (adapted from Pfaff 1979: 299)	176

List of tables

7.1	Verb particle combinations in Sranan–Dutch code-mixing (based on Bolle 1994: 93)	187
7.2	Detailed analysis of occurrences of doubling with Spanish verbs that occur in four <i>waynos</i> or more	190
7.3	Non-loans contrasted with low and high frequency loans; independent occurrences taken together with occurrences in first position of a doublet	190
7.4	The constituency of mixed verbal compounds in Sarnami (based on Kishna 1979)	199
7.5	The form of the verb in mixed verbal compounds in Sarnami (based on Kishna 1979)	201
7.6	The form of mixed verbal compounds involving a verb + particle combination in Sarnami (based on Kishna 1979)	202
7.7	The constituency of mixed verbal compounds in American Greek (based on Seaman 1972)	212
7.8	The use of the Turkish verb <i>yap-</i> ‘do’ in three generations of Turkish migrants in the Netherlands (based on data in Backus 1996: 192, table 5.1)	213
7.9	Summary of the arguments given so far, and their implication for the nominalization, inflection-carrier, and adjunction analyses	218
7.10	Classification of the bilingual verbs in the different settings in terms of the notions insertion (finite, infinitive, compound), alternation, and congruent lexicalization	220
8.1	Diagnostic features of the three patterns of code-mixing	230
8.2	Relative proportion of different categories involved in mixes from Spanish to English in Pfaff (1979) and Poplack (1980) (discounting 747 single noun mixes in Pfaff’s table 5)	232
8.3	A quantitative survey of the sample sentences cited in the major sources for Spanish/English code-mixing in the United States	233
8.4	Code-mixes in Spanish/English in New York (based on Poplack 1980: 602), but with the percentages calculated across the switch directions rather than for the categories switched in one direction, as in the original table	235
8.5	Directionality of multi-word switches in Moroccan Arabic (MA)/Dutch (Du) code-switching (based on Nortier’s table 7.4, 1990: 126)	236
8.6	Directionality of single word switches in Moroccan Arabic	

List of tables

(MA)/Dutch (Du) code-switching (based on Nortier's table 7.15, 1990: 141)	237
8.7 Features of the corpus of Moroccan Arabic–Dutch switching (based on Nortier 1990)	238
8.8 Single-word mixes in the Nairobi corpus (based on Myers-Scotton 1993b)	239
8.9 EL islands (multi-word mixes) in the Nairobi corpus (based on Myers-Scotton 1993b)	239
9.1 The relation of the three mixing strategies to the processes of contact-induced language change	274

I

The study of code-mixing

This book is about intra-sentential code-mixing and how it can help us understand language interaction as the result of contact, yielding a new perspective on central aspects of the human linguistic capacity. The question discussed here is: how can a bilingual speaker combine elements from two languages when processing mixed sentences? I am using the term **code-mixing** to refer to all cases where lexical items and grammatical features from two languages appear in one sentence. The more commonly used term **code-switching** will be reserved for the rapid succession of several languages in a single speech event, for reasons which will be made clear. However, sometimes the terms **switch**, **switch point**, or **switching** will be used informally while referring to the cooccurrence of fragments from different languages in a sentence. Of course, it will also be necessary to separate cases of code-mixing from **lexical borrowing**. The term **language interaction** will be used occasionally as a very general cover term for different, frequently highly innovative, results of language contact, both involving lexical items (as in code-mixing) and otherwise (e.g. phonological or syntactic interference).

In most models portraying the functioning of the speaker/listener, pictures we carry in our minds or see portrayed in a textbook, a single grammar and a single lexicon are embedded in the network of relations that constitutes the model. This is so commonplace that the essential enrichment of having several grammars and lexicons participate in it at the same time is often seen as a threat, a disruption, a malady. This is particularly the case in the structuralist tradition in linguistics. Ronjat (1913) and Leopold (1939–1949) formed the basis for the single parent/single language approach to bilingual child rearing – bilingualism in the family is ok, but it should remain tidy. Weinreich (1953: 73) thought that intra-sentential code-mixing was a sign of lack of bilingual proficiency and interference. An ideal bilingual ‘switches from one language according to appropriate changes in the speech situation (interlocutor, topics, etc.) but not in an unchanged speech situation and certainly not within a single sentence.’ A growing number of studies have

Bilingual speech

demonstrated, however, that many bilinguals will produce mixed sentences in ordinary conversations. What is interesting to me, as it has been to many others in recent years, is that such sentences are produced with great ease and complete fluidity. Indeed, for some speakers it is the unmarked code in certain circumstances (Myers-Scotton 1993a). Neither does it reflect limited proficiency in either of the languages involved. Rather, speakers who code-mix fluently and easily tend to be quite proficient bilinguals (Poplack 1980, Nortier 1990). Finally, we cannot assume either that it is word-finding difficulties or specific cultural pressures that lead to the mixture (even if language contact itself is culturally conditioned). Often, the element introduced corresponds to a household word.

In the last fifteen years, a large number of studies have appeared in which specific cases of intra-sentential code-mixing are analysed from a grammatical perspective. These cases involve a variety of language pairs, social settings, and speaker types. It is found that intra-sentential code-mixes are not distributed randomly in the sentence, but rather occur at specific points. Where much less agreement has been reached is with respect to general properties of the process.

This book is an attempt to present a general account of the very complex intra-sentential code-mixing phenomena that have been discovered. By now the amount of material collected for different language pairs is both diverse and substantial, and it is time to attempt a first synthesis. Rather than introducing one single data set, I will try to integrate the results of a great many different studies, some still unpublished. The present work is grounded both in structural linguistics and in sociolinguistics. Many of the characteristics of the mixing patterns are determined by the structural features of language; I will adopt the general tools and concepts of generative grammar in accounting for these (while trying to stay clear of highly specific formalisms and analyses). Structural analysis along generative lines will be combined with quantitative analysis as in the work of Labov and Sankoff, and comparative typological work. Occasionally, I will try to relate my interpretations to notions from psycholinguistics such as activation and processing. I will only infrequently have recourse to pragmatic and conversational analysis, partly because of my own lack of expertise, and partly because the wide-ranging comparative approach I am adopting here necessarily relies on data gathered by others less suited for detailed textual analysis, and often taking the form of isolated mixed sentences and tables.

The work reported on here could be considered to represent a taxonomic phase in the discipline, an attempt to tie together a set of intermediary results rather than giving a conclusive account. I feel the results from current studies

are so diverse that some tidying up is called for. Although the focus of the present work is grammar, it does not mean attention will not be given to the crucial role of psycholinguistic and sociolinguistic factors influencing code-mixing, such as degree of bilingual proficiency, mode of bilingual processing, political balance between the languages, language attitudes, and type of interactive setting. However, these factors are considered in so far as they are related to or manifest themselves in the grammatical patterns of code-mixing encountered. Indeed, any synthesis at present must depart from the enormous variation in code-mixing patterns encountered, variation due to language typological factors in addition to sociolinguistic and psycholinguistic factors such as those mentioned.

I do not propose a single 'model' of code-mixing, since I do not think there is such a model, apart from the general models provided by grammatical theory and language processing. The challenge is to account for the patterns found in terms of general properties of grammar. Notice that only in this way can the phenomena of code-mixing help refine our perspective on general grammatical theory. If there were a special and separate theory of code-mixing, it might well be less relevant to general theoretical concerns.

Different processes

The patterns of intra-sentential code-mixing found are often rather different from one another. Much of the confusion in the field appears to arise from the fact that several distinct processes are at work:

- **insertion** of material (lexical items or entire constituents) from one language into a structure from the other language.
- **alternation** between structures from languages
- **congruent lexicalization** of material from different lexical inventories into a shared grammatical structure.

These three basic processes are constrained by different structural conditions, and are operant to a different extent and in different ways in specific bilingual settings. This produces much of the variation in mixing patterns encountered. The three processes correspond to dominant models for code mixing that have been proposed.

Approaches that depart from the notion of **insertion** (associated with Myers-Scotton 1993b) view the constraints in terms of the structural properties of some base or matrix structure. Here the process of code-mixing is conceived as something akin to borrowing: the insertion of an alien lexical or phrasal category into a given structure. The difference would simply be the size and type of element inserted, e.g. noun versus noun phrase.

Bilingual speech

Approaches departing from **alternation** (associated with Poplack 1980) view the constraints on mixing in terms of the compatibility or equivalence of the languages involved at the switch point. In this perspective code-mixing is akin to the switching of codes between turns or utterances. This is the reason I avoid using the term **code-switching** for the general process of mixing. Switching is only an appropriate term for the alternational type of mixing. The term code-switching is less neutral in two ways: as a term it already suggests something like alternation (as opposed to insertion), and it separates code-mixing too strongly from phenomena of borrowing and interference.

The distinction I make here between alternation and insertion corresponds to Auer's distinction between code-switching and transfer (1995: 126). Some authors have used the term 'switching' for language interaction between clauses, and 'mixing' for intra-clausal phenomena. This distinction parallels my distinction between alternation and insertion, but does not coincide with it, since in my framework alternation often takes place within the clause as well.

The notion of **congruent lexicalization** underlies the study of style shifting and dialect/standard variation, as in the work of Labov (1972) and Trudgill (1986), rather than bilingual language use proper. The exception is the bilingual research by Michael Clyne (1967) on German and Dutch immigrants in Australia. This comes closest to an approach to bilingual language use from the perspective of congruent lexicalization.

In this book I am claiming that these different models or approaches in fact correspond to different phenomena: there is alternation between languages, insertion into a matrix or base language, and congruent lexicalization, in the code-mixing data reported in the literature. In chapters 3, 4, and 5 criteria are proposed, both structural and quantitative, for giving substance to the three-way distinction.

In some cases, a single constituent is **inserted** into a frame provided for by the matrix language:

- (1) kalau dong bukan bikin dong bukan bikin
when they always make they always make
voor acht personen dek orang cuma nganga dong makan
for eight persons and then people only look they eat
'When they [cook], it is always for eight people, and then they only look at it, they eat . . .'

(Moluccan Malay/Dutch; Huwaë 1992)

While in (1) this is an entire Dutch prepositional phrase inserted into a Moluccan Malay sentence, in (2) it is a single English verb stem used in a complex Navaho verbal structure:

- (2) na'iish-crash lá
1sg:pass out-crash EMPH
'I am about to pass out.'
(Navaho/English; Canfield 1980: 219)

In (3) the temporal expression *por dos días* is clearly related to the verb *anduve*, encapsulating the inserted *in a state of shock*:

- (3) Yo anduve *in a state of shock* por dos días.
'I walked in a state of shock for two days.'
(Spanish/English; Pfaff 1979: 296)

With insertion, there is embedding. The English prepositional phrase is inserted into an overall Spanish structure. Insertion is akin to (spontaneous) lexical borrowing, which is limited to one lexical unit. There is considerable variation in what is or can be inserted: in some languages this consists mostly of adverbial phrases, in others mostly single nouns, and in yet others again determiner + noun combinations. Insertion and the distinction between code-mixing and borrowing are taken up again in chapter 3.

In other cases, it seems that halfway through the sentence, one language is replaced by the other. The two languages **alternate**:

- (4) maar 't hoeft niet *li-'anna ida šeft ana . . .*
but it need not for when I-see I
'but it need not be, for when I see, I . . .'
(Moroccan Arabic/Dutch; Nortier 1990: 126)
- (5) Les femmes et le vin, *ne ponimayu.*
'Women and wine, I don't understand.'
(French/Russian; Timm 1978: 312)
- (6) Andale pues *and do come again.*
'That's all right then, and do come again.'
(Spanish/English; Gumperz and Hernández-Chavez 1971: 118)

In the case of alternation, there is a true switch from one language to the other, involving both grammar and lexicon. Thus in (6) there is no reason to assume that the Spanish first segment is embedded in the English second segment or vice versa. Alternation is just a special case of code-switching, as it takes place between utterances in a turn or between turns. In chapter 4 alternation is studied in more detail.

In a third set of cases, it appears that there is a largely (but not necessarily completely) shared structure, lexicalized by elements from either language, **congruent lexicalization**. Consider the following examples:

- (7) Weet jij [*waar*] Jenny is?
'Do you know where Jenny is?' (Dutch: waar Jenny is)
(English/Dutch; Crama and van Gelderen 1984)

The sequence *where Jenny is* could as easily be English in structure as Dutch. Furthermore *where* is close to Dutch *waar* (particularly when pronounced by bilinguals), *Jenny* is a name in both languages, and *is* is homophonous.

Bilingual speech

A similar example is:

- (8) En de partij dy't hy derby blaasde, (Frisian)
en de partij die hij erbij blies (Dutch)
And the part that he thereby blew
is net [**foar** *herhaling vatbaar*]. (Frisian)
is niet voor herhaling vatbaar (Dutch)
is not for repetition handable.
'And the song he sang then is not fit to be repeated.'

(Frisian/Dutch; Wolf 1995: 12)

Here, Frisian *foar* 'for' is sufficiently similar to Dutch *voor* 'for' to be an ambiguous switchpoint; Dutch *herhaling vatbaar* is not a constituent, but two words that form an idiom together with *voor*.

While English/Dutch and Frisian/Dutch are two closely related language pairs with many cognates, we may find something similar to these examples in the English/Spanish material analysed by Poplack (1980) as well:

- (9) (A) Why make Carol *sentarse atrás* (B) *pa'que* everybody
sit at the back so that
has to move (C) *pa'que se salga*.
so that [she] may get out. (Spanish/English; Poplack 1980: 589)

Here sentence fragment (B) is a complement to (A), and (C) is a complement to (B). Notice that the first Spanish fragment here contains both a verb phrase, *sentarse atrás* and a purposive complementizer, *pa'que*. There is no particular grammatical relation between the two English fragments nor between the Spanish ones. The example could perhaps be analysed as back-fire insertions within insertions. However, this is rather counter-intuitive, both because the switched fragments are not unique constituents and because they do not appear to obey rules specific to the supposed matrix constituent, but rather rules common to both languages.

Consider a final example:

- (10) Bueno, *in other words*, el *flight* [que sale de Chicago *around three o'clock*].
'Good, in other words, the flight that leaves from Chicago around three
o'clock.' (Spanish/English; Pfaff 1976: 250)

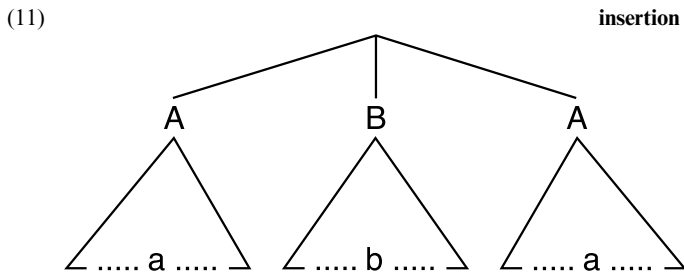
In (10) *que sale de Chicago* 'that leaves Chicago' or even *el flight que sale de Chicago* (assuming *Chicago* to be part of the Spanish stretch for the sake of the argument – in fact it may be the trigger for the subsequent switch to English) is a constituent, but not a unique one, since it also includes the English fragment *around three o'clock*.

The term congruent lexicalization refers to a situation where the two languages share a grammatical structure which can be filled lexically with elements from either language. The mixing of English and Spanish could be

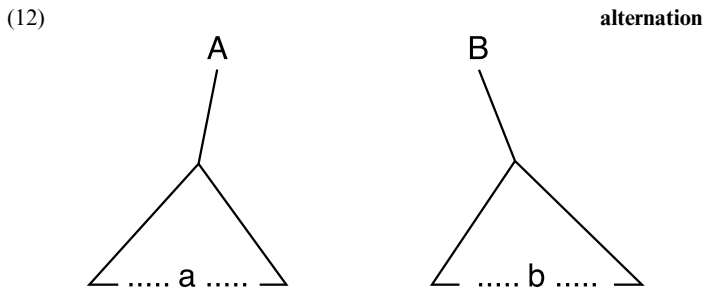
interpreted as a combination of alternations and insertions, but the going back and forth suggests that there may be more going on, and that the elements from the two languages are inserted, as constituents or as words, into a shared structure. In this perspective, congruent lexicalization is akin to style or register shifting and monolingual linguistic variation. The latter would be the limiting case of congruent lexicalization.

I want to explore these three separate patterns of intra-sentential code-mixing and study them through the systematic exploration of bilingual corpora, in addition to the detailed structural analysis of individual examples (van Hout and Muysken 1995).

The structural interpretation of these notions is as follows. Consider the following trees, in which *A*, *B* are language labels for non-terminal nodes (i.e. fictitious markers identifying entire constituents as belonging to one language), and *a*, *b* are labels for terminal, i.e. lexical, nodes, indicating that the words chosen are from a particular language.



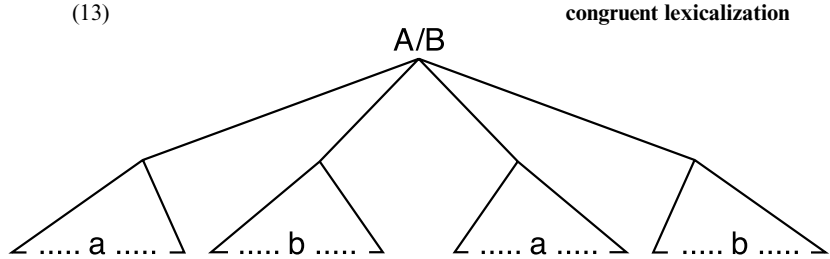
In this situation, a single constituent *B* (with words *b* from the same language) is inserted into a structure defined by language *A*, with words *a* from that language.



In this situation, a constituent from language *A* (with words from the same language) is followed by a constituent from language *B* (with words from

Bilingual speech

that language). The language of the constituent dominating *A* and *B* is unspecified.



Finally, in (13) the grammatical structure is shared by languages *A* and *B*, and words from both languages *a* and *b* are inserted more or less randomly.

Having presented the three processes, insertion, alternation, and congruent lexicalization, I will try to suggest a number of diagnostic features which may be used to distinguish these three patterns in chapters 3, 4, and 5. There are a number of criteria I would like to consider, and I will illustrate their application with concrete cases. The criteria are rarely knock-down criteria by themselves, but should be used conjointly to characterize a sentence or a bilingual speech sample as a case of alternation, insertion, or congruent lexicalization.

In addition to the **structural** interpretation of the three patterns, in terms of labels in tree configurations, there can also be a **psycholinguistic** and a **sociolinguistic** one. The **psycholinguistic** interpretation of the three-way distinction made here could be in terms of different degrees of activation of components of both languages in speech production. In the case of alternation, activation would shift from one language to another, and in the case of insertion, activation in one language would be temporarily diminished. For congruent lexicalization, the two languages partially share their processing systems. Psycholinguistic factors determining the choice between these different processes include bilingual proficiency, level of monitoring in the two languages, the triggering of a particular language by specific items and the degree of separateness of storage and access systems.

The interpretation of the three patterns can also be **sociolinguistic**, in terms of bilingual strategies (an example would be Sankoff, Poplack, and Vanniarajan's (1990) distinction between equivalence and insertion). The sociolinguistic embedding of these three patterns, i.e. their use as bilingual strategies, can be described as follows. The process of alternation is particularly frequent in stable bilingual communities with a tradition of language separation, but occurs in many other communities as well. It is a frequent

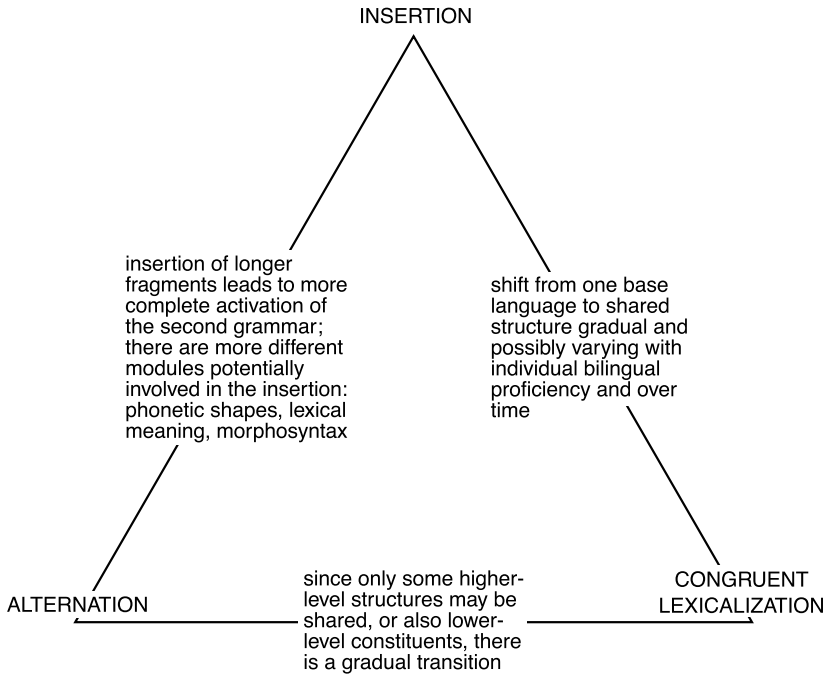


Figure 1.1 Schematic representation of the three main styles of code-mixing and transitions between them

and structurally little intrusive type of code-mixing. Insertion is frequent in colonial settings and recent migrant communities, where there is a considerable asymmetry in the speakers' proficiency in the two languages. A language dominance shift, e.g. between the first and third generation in an immigrant setting, may be reflected in a shift in directionality of the insertion of elements: from insertion into the language of the country of origin to the presence of originally native items in the language of the host country. Congruent lexicalization may be particularly associated with second generation migrant groups, dialect/standard and post-creole continua, and bilingual speakers of closely related languages with roughly equal prestige and no tradition of overt language separation.

The three types of code-mixing can be conveniently viewed in terms of a triangle, an image to which I return in chapter 8 when I discuss the various factors that help determine which type of mixing occurs in a specific setting. The differences between the three types are gradual rather than absolute, as can be seen in figure 1.1.

Between **insertion** and **alternation** there is a transition zone since insertion

Bilingual speech

of longer fragments leads to increasingly more complete activation of the second grammar. There are different modules potentially involved in the insertion: phonetic shapes, lexical meaning, morphosyntax, which involve activation of the second grammar to different degrees.

Since only some higher-level structures may be shared between languages, there is a gradual transition as well between **alternation** and **congruent lexicalization**: alternation when only the top node (the sentence node) is shared, and congruent lexicalization when all or most nodes are shared between the two languages.

Finally, there can be a gradual shift from one base language to shared structure and on to the other base language, possibly varying with individual bilingual proficiency and over time. This implies particularly that in many immigrant communities, **insertion** of new items and expressions into the home language can evolve into **congruent lexicalization** and then possibly into **alternation** (with set phrases and expressions from the ethnic language interspersed in the new language).

The literature abounds both with proposals for various specific constraints on code-mixing, and with claims that general constraints do not hold. Romaine (1989, 1995) takes an intermediate position, namely that none of the constraints covers all cases. A similar position is taken by Clyne (1987). I want to argue against too much relativism, however, both because it is overly pessimistic of the relevance of linguistic structure, and because I believe it only portrays a limited picture of the often quite regular array of code-mixing patterns to be found. Rather, a particular constraint can only be assumed to hold for a specific type of mixing.

The role of a dominant research experience in shaping one's vision: the development of code-mixing research

It has been said somewhat maliciously that an Indo-Europeanist will tend to make reconstructed Proto-Indo-European look like the language that she or he thoroughly studied early on. If the scholar starts out with Albanian, her or his version of the proto-language will have many features of Albanian, and so on. This subjectivism (if the observation is at all valid) may seem reprehensible, but I think there are many ways in which research that one does early on shapes one's vision of a certain domain. This does not mean one is condemned to the one perspective, but rather that this perspective is the frame of reference onto which later ideas are often grafted. In this way, research on code-mixing has been shaped by the language pairs encountered by researchers, and the key notions were introduced one by one on the basis of the data encountered.

Two scholars started with code-mixing research in the 1960s: Meri Lehtinen and Michael Clyne. Thirty years ago Meri Lehtinen wrote, in her study of the recorded Finnish–English bilingual corpus of one speaker:

In order for any intra-sentence code-switching to be possible at all, there must exist in the two languages some constructions which are in some sense similar, so that certain syntactic items from each language are equivalent to each other in specific ways. Further reflection, supported by an examination of the corpus, shows that the similarities must exist in what is known as the ‘surface grammar’ of sentences (1966: 153)

And further on: ‘It would seem, then, that switching at words which belong to a closed class is not allowed by the code-switching rules, except in cases where such a switch is forced by structural considerations’ (1966: 177). The three notions of **syntactic equivalence**, **surface linear order**, and **closed class items** are still at the core of our view of intra-sentential code-mixing. In this sense, everything written since is a comment on and elaboration of Lehtinen’s work. Where we have progressed enormously since Lehtinen is in the empirical scope of the research and in the level of explanation advanced. Explaining why these notions are important in accounting for bilingual speech also makes it possible to study the relation between them and their relative weight in an explanatory model.

Earlier work by Michael Clyne has helped shape the view of code-mixing as something relatively unconstrained, like congruent lexicalization. Clyne has worked in Australia, particularly with the German and Dutch immigrant communities. His best-known books in this area are *Transference and triggering* (1967) and *Perspectives on language contact* (1972). In his work he places the phenomenon of code-mixing in the context of a complex set of other language contact processes: **lexical triggering**, **transference**, and **convergence**. The perceived similarities between the languages bilinguals speak facilitate code-mixing, and in the bilingual setting the languages will tend to converge. The use of a word from another language may easily trigger other material from that language, either in anticipation of that word or subsequently. The picture dominant in Clyne’s work then is neither that of alternation nor of insertion, but of congruent lexicalization. Consider examples such as:

- (14) Wan ik komt *home from school*.
‘When I come home from school.’ (English/Dutch; Clyne 1987: 759)

Here English *when* appears as a Dutch-like word *wan* (cf. Du *wanneer*), the Dutch word order is adapted to English (cf. Du *wanneer ik thuis* (‘home’) *kom*), and the inflected form of the verb is not correct. Similarly:

Bilingual speech

- (15) That's what *Papschi mein*-s to say.
That's what Papschi means to say. (German/English; Clyne 1987: 756)

The use of a German name, *Papschi*, triggers a German verb *mein*, which is quite close semantically to English *mean*, to be sure, and inflected with English third person *-s*.

Recognizing the sociolinguistically determined nature of code-mixing, a considerable part of the research since 1970 has focussed on the syntactic properties of code-mixing: where in the sentence do we find it, and when is it impossible? In other words, what are the constraints on code-mixing? This research has undergone three stages (Appel and Muysken 1987; Bhatia and Ritchie 1996): (a) an early stage in which grammatical constraints specific to particular constructions were focussed on; (b) a stage which has produced the classical studies in which universal constraints on code-mixing were explored, from around 1980 onward; (c) the present stage, which may be characterized by the search for new perspectives: what alternative mixing strategies are there and are constraints perhaps relative to a particular strategy?

Particular grammatical constraints

Most of the code-mixing studies from the 1970s drew on Spanish–English data recorded from conversations by Mexican Americans and Puerto Ricans, and proposed construction-specific constraints. Gumperz and Hernández-Chavez (1971) noted that mixing was easily possible in some contexts, but not so much in others. Contexts allowing a switch included:

- (16) Between a head noun and a relative clause:
those friends are friends from Mexico *que tienen chamaquitos* (that have little children) (Spanish/English; Gumperz and Hernández-Chavez 1971: 118)
- (17) Between a subject and a predicate in a copular construction:
An' my uncle Sam es *el más agabachado*. (is the most Americanized)
(Spanish/English; Gumperz and Hernández-Chavez 1971: 119)

Switches as in (18), with the verb in English, are not allowed, however:

- (18) * que *have* chamaquitos

In a more systematic treatment Timm (1975) proposed the following restrictions:

- (19) Subject and object pronouns must be in the same language as the main verb:
* *Yo* (I) went.
* Mira (look at) *him*
- (20) An auxiliary and a main verb, or a main verb and an infinitive must be in the same language:

* They want *a venir* (to come).

* ha (he has) *seen*

In the studies of Gumperz and Hernandez-Chavez and Timm just cited two methodologies are combined: the analysis of recorded conversations and grammaticality judgements. The stars in the above examples reflect judgements of bilinguals about possible switches, but that these judgements do not always correspond to actual mixing behaviour is clear when we compare the two observations by Lipski (1978) in (21) and (22) with findings of Pfaff (1979):

(21) It is difficult to switch inside a prepositional phrase:

?? in *la casa* (the house)

(22) It is impossible to switch between the article and the noun:

?? I see the *casa* (house).

Both observations contrast with a large number of cases of precisely these switches found in the corpus analysed by Pfaff (1979). Mixing internal to prepositional phrases (PPs), i.e. of English nouns into Spanish PPs occurs far more often than mixing at PP boundaries. We also find more cases of a switch between the article and the noun than switches between article + noun combinations and the rest of the sentence. Clearly it is difficult if not impossible to rely on judgement data.

The studies of code-mixing carried out in the 1970s provide us with a large body of analysed data, with a number of inductive generalizations, and with insights into the type of constraints on code-mixing to expect. An overall theoretical perspective was lacking, however, and this is what the studies of the early 1980s have tried to contribute.

Poplack's Equivalence and Free Morpheme constraints

Although many of the ideas voiced in the current code-mixing literature were informally proposed earlier, the major impetus for a more systematic exploration of bilingual data came with Poplack (1980). I will take that date and reference as the starting point for my discussion of different models.

In her work on Spanish/English code-mixing in the Puerto Rican community (1980), Poplack discovered that code-mixing occurred largely at sites of equivalent constituent order. She is the principal exponent of the alternation perspective, and stresses the importance of **linear equivalence** between the languages involved at the point of the switch.

The order of sentence constituents immediately adjacent to and on both sides of the switch point must be grammatical with respect to both languages involved simultaneously . . . The local co-grammaticality or equivalence of the two languages in the vicinity of the switch holds as

Bilingual speech

long as the order of any two sentence elements, one before and one after the switch point, is not excluded in either language.

(Sankoff and Poplack 1981: 5)

Underlying this constraint is a particular definition of code-switching, which embodies the idea of alternation, phrased by Poplack (1993) as follows: 'Code-switching is the **juxtaposition** [emphasis SP] of sentences or sentence fragments, each of which is internally consistent with the morphological and syntactic (and optionally, phonological) rules of the language of its provenance.' An example of Spanish/English code-mixing illustrating this juxtaposition is:

- (23) (A) *Se me hace que* (B) I have to respect her (C) *porque 'ta* . . . older.
'It appears to me that I have to respect her because [she] is . . . older.'

(Spanish/English; Poplack 1980: 591)

Fragment (B) is a complement to (A), and (C) modifies (B). Notice that *porque 'ta* 'because [she] is' does not form a unique constituent, excluding other elements, in this case *older*. It is clear that this type of data (see also example (9) above) cannot be handled very well in a model which takes insertion into a matrix and a dependency relation between matrix and inserted material as its primes. Rather, it has led to the idea that order equivalence across the switch point is what constrains code-mixing here.

The constituent order equivalence constraint is illustrated with the following example:

- (24) I told **him** that so that he would bring **it** fast
 | X | | | X |
(Yo) **le** dije eso pa' que (él) **la** trajera rápido

(Spanish/English; Poplack 1980: 586)

Mixing is impossible where there is a difference in word order between Spanish and English. Here this is particularly the case around the object pronouns.

In her 1980 paper, Poplack proposed an additional principle, the **Free Morpheme Constraint**, which holds that:

- (25) Codes may be switched after any constituent in discourse provided that constituent is not a bound morpheme. (Poplack 1980: 585–6)

Presumably switches both **before** and **after** a bound morpheme are prohibited by this constraint.

Soon, a number of criticisms were raised with respect to both the Equivalence Constraint and the Free Morpheme Constraint, often involving typologically more different language pairs, where code-mixing appeared to violate both constraints at the same time. Partly in response to this, Poplack has developed a more elaborate typology of code-mixing phenomena in later

work, always taking as the starting point the issue of whether a given code-mixing pattern conforms to the equivalence constraint. She has argued that many cases that appear to violate both the Free Morpheme Constraint and the Equivalence Constraint were actually inflected (**nonce**) borrowings. Other cases violating the Equivalence Constraint are analysed as cases of constituent insertion. In subsequent papers, some written in collaboration with David Sankoff and others, Poplack developed methods and criteria for characterizing the contrasts between nonce borrowing and code-mixing, and equivalence-based mixing versus constituent-insertion. These studies have met with some criticism, and will be discussed in detail in chapter 3 on insertion. I should say that personally I am sympathetic to Poplack's attempt to separate the different types of mixing, even though I am not in agreement with the precise boundaries she draws between these types.

The typology of language contact phenomena developed by Poplack, Sankoff, and co-workers has focussed on four types of mixing so far:

- (26) code-switching under equivalence
 - (nonce) borrowing
 - constituent-insertion
 - flagged switching

Flagged switching will be discussed in chapter 4 as indicative of alternation.

The Matrix Language Frame model

The data on which the perspectives taken by scholars such as Clyne and Poplack were based contrast rather sharply with the cases of Swahili–English mixing that have been the basis for Myers-Scotton's work, which exemplifies the insertion approach (1993b). Carol Myers-Scotton is best known for her research on Swahili–English bilingualism in eastern Africa, which she has approached from a number of perspectives: strategies of neutrality, and code-mixing as a marked or unmarked choice, and a comprehensive psycholinguistically embedded linguistic model for intra-sentential code-mixing. Myers-Scotton (1993a: 4) gives the following definition: 'Code-switching is the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation.' This definition, which differs from that of Poplack cited before, is in line with the author's structural work and fits much of the African material discussed (characterized by insertions). It makes it necessary, however, to assume a going back and forth between different matrix languages where e.g. Spanish–English code-mixing in New York is discussed.

Bilingual speech

Neither is it clear that the central notion of ‘unmarked code-switching’ requires the concept of a matrix language.

Consider some of the cases that form the basis of Myers-Scotton’s analysis, which are representative of the data reported on in her work:

- (27) Na kweli, hata mimi si-ko *sure* lakini n-a-*suspect* i-ta-kuwa *week* kesho.
‘Well, even I am not sure, but I suspect it will be next week.’

(Swahili/English; Myers-Scotton 1993b: 81)

Here the elements *sure*, *suspect*, and *week* are single elements inserted into a Swahili sentence.

The proto-typical type of example cited by Myers-Scotton corresponds directly to her view of the processes of code-mixing as a whole. Nishimura (1986: 126) also notes the tendency for researchers working on typologically similar languages such as Spanish or German and English to adopt symmetrical models (involving alternation or congruent lexicalization), and researchers working on typologically dissimilar language pairs such as Marathi or Swahili and English to adopt asymmetrical, insertional, models.

It is obvious that the kind of material analysed by Clyne gives rise to a quite different perspective on the phenomenon of code-mixing than the data from Spanish/English and Swahili/English that were the starting point of Poplack’s and Myers-Scotton’s research experiences.

For code-mixing of the insertional type a theoretical framework is provided by Myers-Scotton’s Matrix Language Frame model (1993b, 1995), in which the matrix language constituent order and matrix language functional categories are assumed to dominate a clause. The model proposed, the Matrix Language Frame model, crucially incorporates the idea that there is an asymmetrical relation between the **matrix** and the **embedded** language in the mixing situation. Furthermore, **content** and **function** morphemes behave differently in Myers-Scotton’s model: the former can be inserted into mixed constituents, when congruent with the matrix language categories, while the latter cannot. Finally, no essential difference is made between mixing and borrowing at the level of morphosyntactic integration, as it is in Poplack’s work. The model proposed rests on the assumption that mixed sentences have an identifiable base or matrix language (ML), something that may or may not hold for individual bilingual corpora. There is always an asymmetry between the ML and the embedded language (EL).

I will primarily rely on the presentation of Myers-Scotton and Jake (1995), also citing Myers-Scotton (1993b) where necessary. The model makes the following claims:

- (a) The ML determines the order of the elements in mixed (ML + EL) constituents (**Morpheme Order Principle**; Myers-Scotton and Jake 1995: 983):

In ML + EL constituents consisting of singly occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relationships) will be that of the ML.

(b) There is a fundamental difference in distribution of functional elements and content words in mixed sentences: functional elements in mixed ML + EL fragments can only be drawn from the ML. The ML provides the ‘system morphemes’ (functional categories) in such constituents (the **System Morpheme Principle**; Myers-Scotton and Jake 1995: 983):

In ML + EL constituents, all system morphemes that have grammatical relations external to their head constituent (i.e. participate in the sentence’s thematic role grid) will come from the ML.

(c) In mixed constituents only certain EL content morphemes may occur (the **Blocking Hypothesis**; Myers-Scotton 1993b: 120):

In ML + EL constituents, a blocking filter blocks any EL content morpheme which is not congruent with the ML with respect to three levels of abstraction regarding subcategorization.

Here **congruence** refers to ‘a match between the ML and the EL at the lemma level with respect to linguistically relevant features’ (Myers-Scotton and Jake 1995: 985). The three levels of abstraction are: having the same status in both languages, taking or assigning the same thematic roles, and having equivalent pragmatic or discourse functions. Researchers have often stressed that there needs to be some kind of categorial equivalence or congruence (Myers-Scotton’s term) between the constituent inserted and the matrix language node into which it is inserted (e.g. Sebba 1998). The question is how to define this notion across languages, and whether strict identity of features is assumed, only compatibility, or translation equivalence. I return to this in the next chapter.

The two languages are separately processed in units called islands. Three types of constituents are listed: EL Islands, ML Islands, and Mixed EL + ML constituents. In the Matrix Language Frame model insertions correspond to mixed EL + ML constituents, alternations to EL islands combined with ML islands, and congruent lexicalization is akin to a notion that Myers-Scotton is developing in as yet unpublished work on the possibility of a ‘composite ML’.

A recent development in the work of Myers-Scotton and Jake (e.g. Jake and Myers-Scotton 1997), building on Myers-Scotton and Jake (1995) has been the attempt to explain the incidence and patterning of code-mixing through the notion of compromise strategies, strategies meant to avoid a clash in congruence between the properties of an inserted lemma (mental

Bilingual speech

representation of a lexical item) and properties of the matrix language. Several strategies are suggested, including: the creation of larger EL islands encapsulating the non-congruent lexical item, and the use of bare forms. These strategies run parallel to the approach taken here and outlined at the end of this chapter. It is an empirical question, however, whether all relevant examples can be explained in terms of the avoidance of a lexical congruence clash. Another development is the further refinement of the content word/system morpheme distinction, e.g. in Myers-Scotton (1999). This will be taken up in more detail in chapter 6.

Myers-Scotton's work has been criticized on a number of counts: the notion of ML is often too rigid, the definition of system morphemes is problematic, it is difficult to find an appropriate definition of congruence, and the psycholinguistic processing model assumed is not fully explicit. Nonetheless, Myers-Scotton has drawn together psycholinguistic, sociolinguistic, and structural perspectives on code-mixing for the first time, and thus brought its study to a deeper explanatory level.

System morphemes (= functional elements) will be discussed in chapter 6. What is the role of functional categories in code-mixing processes? Do they determine the matrix language frame? Does (lack of) equivalence between functional categories in different languages play an important role? How do we define and identify functional elements?

One of the sources of inspiration for the Matrix Language Frame model lies in the work of Joshi (1985). Joshi has come up with an asymmetrical model on the basis of data from Marathi/English code-mixing and considerations from the mathematical theory of syntactic parsing. Crucial to Joshi's work is the notion of closed-class item, which cannot be switched. Doron (1983) has expanded his model, arguing on the basis of considerations involving left-to-right parsing that the first word of a sentence or a constituent determines the host or base language. Properties of the host language determine, in Joshi's perspective, whether mixing is possible or not, including the selection of closed class items or function words.

Another model closely related to the Matrix Language Frame model is presented in Azuma's work (1993). Departing from Garrett's (1982) speech production model, Azuma formulates the Frame-Content Hypothesis, in which the frame-building stage, 'where closed-class items are accessed and retrieved' (1993: 1072) precedes a content-word insertion stage. Though much less elaborate than Myers-Scotton's work, Azuma's model seems to be making roughly the same predictions.

Myers-Scotton's original model has been criticized on a number of grounds: the definition of matrix language, the distinction between system

and content morphemes, and the relation between code-mixing and language processing. I will address these criticisms in more detail in chapters 3, 6, and 9, respectively.

Dependency, coherence, and sentence organization

I should perhaps make clear how my own research in the area of language contact started. In 1977 I had already been doing eight months of fieldwork on the Andean Amerindian language of Quechua in a community in central Ecuador, where both Quechua and Spanish were spoken. In the middle of the night I heard my hosts at the time speak yet a third language among themselves. Upon enquiry the next morning I found out that this language, which had sounded entirely strange to me, was really a highly innovative mixture of Quechua and Spanish. It was often referred to as Media Lengua ‘half(way) language’, or Utila Ingiru ‘little Inca-ese (Quechua)’ (Muysken 1981a, 1996). In the following example the Media Lengua (MeL) original is given along with both its Quechua (Qe) and Spanish (Sp) equivalents.

- (28) MeL *Chicha-da-ga xora-mi irbi-chi-ndu, ahi-munda-mi chicha-AC-TO*
corn-AF boil-CAU-SUB there-ABL-AF
Qe *Aswa-da-ga sara-mi yanu-sha, chay-munda-mi*
Sp *Chicha, haciendo hervir jora, después*
MeL *sirni-nchi, ahi-munda-ga dushi-da poni-nchi.*
strain-1pl there-ABL-TO sweet-AC put-1pl
Qe *shushu-nchi chay-munda-ga mishki-da chura-nchi.*
Sp *la cernimos, y después la ponemos dulce.*
‘As to chicha, having boiled corn first we strain it and then we put in
sugar.’ (Media Lengua; Muysken, fieldwork data)

An inspection of this recorded sentence and its equivalents will reveal that all lexical bases in Media Lengua are Spanish, the affixes all Quechua (with the exception of the gerundive marker *-ndu*, <Sp *-ndo*), and the general word order and syntax Quechua.

A number of the questions raised by Media Lengua turn out to be the same as the questions that started to intrigue me about code-mixing in subsequent years. These include the role of morphology in language mixing; the importance of syntagmatic and paradigmatic coherence; the interaction of the lexicon and the morpho-syntax; the role of typological differences between the languages involved; and finally, simultaneous rather than sequential operation of properties of the languages involved in the mix.

The work led to a third type of approach that stresses **dependency** rather than equivalence, assuming that code-mixing obeys a general constraint of lexical dependency. The basic idea was that there cannot be a mix between

Bilingual speech

two elements if they are lexically dependent on each other. A first implicit statement of this restriction is given in Shaffer (1978), but a more explicit formulation appeared in DiSciullo, Muysken, and Singh (1986), in terms of the **government** model. In this model, the relation between a head and its syntactic environment, as circumscribed by government, was assumed to constrain possible code-mixes.

Proposals similar to the one by DiSciullo, Muysken, and Singh (1986) have been put forward by Bentahila and Davies (1983) and by Klavans (1985). Klavans argues that it is the language of the inflected main verb or the auxiliary of a clause that determines the restrictions on code-mixing in that particular clause, since those elements in some sense constitute the syntactic head of the clause and govern the rest.

Bentahila and Davies (1983) propose that the **subcategorization** properties of a word determine what elements, including elements of another language, may appear within the phrase syntactically headed by that word. The following contrasts illustrate their approach.

- | | | | |
|------|----|----------------------------|-------------------|
| (29) | a. | * <i>cette l xubza</i> | ‘this the loaf’ |
| | | * <i>un l fqi</i> | ‘one the teacher’ |
| | b. | <i>cette xubza</i> | ‘this loaf’ |
| | | <i>un fqi</i> | ‘one teacher’ |
| (30) | a. | * had <i>pain</i> | ‘this loaf’ |
| | | * wahed <i>professeur</i> | ‘a teacher’ |
| | b. | had <i>le pain</i> | ‘this loaf’ |
| | | wahed <i>le professeur</i> | ‘a teacher’ |

(Moroccan Arabic/French; Bentahila and Davies 1983: 109)

The mixes in (29a) and (30a) are ungrammatical, in their view, because the French determiners in (29) subcategorize for a simple noun without the article *l* (as shown in (29b)), and the Arabic determiners in (30) subcategorize for a noun with an article (as shown in (30b)). In neither case is there a violation of the word order of either language. Something like the notion of government is at play: for Bentahila and Davies’ proposal to work, they have to assume that the determiner and the rest of the noun phrase are in the government relation of selection.

Government is a traditional grammatical notion which has received several formulations within the theory of Government and Binding, e.g. as in Chomsky (1981: 164):

- (31) α governs γ in [$\beta \dots \gamma \dots \alpha \dots \gamma \dots$], where:
(i) $\alpha = X^0$
(ii) α and γ are part of the same maximal projection.

Typical cases of government would be case assignment, as in the Latin