

- Wei, L. (2002). 'What do you want me to say?' On the Conversation Analysis approach to bilingual interaction. *Language in Society* 31(2): 159–80.
- (ed.) (2005). Special Issue: Conversational Code-Switching. *Journal of Pragmatics* 37: 275–410.

15 the monolingual bias in bilingualism research, or: why bilingual talk is (still) a challenge for linguistics

peter quer

introduction

For a long time, linguists found it difficult to account for the use of two or more 'languages' within one utterance by the same speaker. It was acknowledged of course (from the nineteenth century onwards, at the latest) that languages can 'borrow' structures from other languages without ever returning them to the 'owners' (to stick to this somewhat problematic metaphorical field). No doubt languages such as German or, even more so, English had massively copied lexical and – to a lesser extent – grammatical elements (above all derivational affixes) from other languages, such as Latin or French. However, these borrowings were exclusively analysed post factum, i.e. after they had become fully incorporated into the borrowing language. Few linguists were interested in languages whose status was unstable and ambiguous; among them was the Austrian Hugo Schuchardt who investigated 'mixed' languages such as creoles and Romani varieties as early as 1884 and came to the conclusion 'dass eine Sprache A ganz allmählich, durch fortgesetzte Mischung, in eine von ihr sehr verschiedene B übergehen kann' [that a language A can transform slowly but steadily, by constant mixture, into a language B which is very different from it]. He continued on a somewhat fatalistic note: 'Für die Beantwortung der Frage aber ob sie an einem bestimmten Entwicklungspunkt noch A oder schon B zu nennen sind, fehlt es uns gänzlich an Kriterien' [However, we would lack all criteria to answer the question whether they can still be called still A or already B at a certain point of development] (1884: 10).

On the level of the individual, the coexistence of two languages only started to become a topic some time later, when linguists became interested in speakers with a foreign 'accent', or in those who 'wrongly' used the grammar of

their second language because their 'mother tongue' had a different structure and the acquisition of the second language had not been fully successful. In contradistinction to borrowing, processes of such 'interference' (as this seemingly mechanistic overlay of one language over the other was aptly called; cf. Weinreich 1953) were considered to be entirely derivative of the two language systems (or their cognitive representations) which interfere with each other in the individual. What was observed in the (imperfect) bilingual was analysed as either a transient or a non-successful case of second language acquisition which left the interfering languages unaffected. Both phenomena (borrowing on the level of the linguistic system and interference on the level of the individual learner) were thereby treated in such a way that they did not question the existence and autonomy of the two languages in contact. Both could be handled within an approach to language which started from the seemingly trivial assumption that there *were* languages such as German, Latin or French which could be neatly and unambiguously separated from one another.

In this chapter I argue that the most profound way in which bilingual talk (code-switching, mixing, and the rest) has challenged linguistics is the fact that its analysis leads to the inevitable conclusion that this assumption cannot be taken for granted. Bilingual talk blurs the line between language A and language B, but also between 'langue' and 'parole', between linguistic systems and their usage, between knowledge and practice. It questions the starting point of linguistics as a whole: in code-switching studies, it turns out that a 'language' cannot be a prime of linguistic analysis (cf. Le Page 1989). The idea that linguistics has as its object self-contained linguistic entities (systems) called languages, and that bilingualism is a derivative of the combination of two of these languages, is of course part and parcel of the nation-state language ideology which dominated European thinking at the time when linguistics established itself as a discipline. The nation-state ideology views a language as a 'natural' reflex of nationhood (and inversely, a common language as a justification for nation-building); each nation has one (and only one) national language. It led to the compartmentalization of linguistic research into the European national philologies which left little room for research on those cases which the model obviously did not fit. More important than disciplinary divisions, however, was the reification of the national languages which this language ideology implied: they were treated as 'natural' facts, despite the enormous purificational efforts most European nations had to invest in their very creation. For the burgeoning new discipline of (structural) linguistics, these languages appeared to be (the only) solid objects. It was forgotten that they showed a high degree of structural regularity and homogeneity precisely as a consequence of the standardization processes which had led to their emergence, and the language ideology which accompanied it. In this framework, bilingualism, if it was considered at all, could only be treated as a form of individual deviance,

from which an individual 'suffered' unless s/he was a 'perfect' bilingual (i.e. double monolingual), and sometimes even then: obviously, if belonging to a nation was identified with speaking its (one) language, then speaking two languages implied belonging to two nations, having two national identities; and how could somebody be both 'German' and 'French', or both 'Japanese' and 'Chinese'?

This benign neglect of bilingual talk lasted until the middle of the last century. The attraction it has developed since then for linguists of all backgrounds neatly mirrors the transformation linguistics underwent as a discipline at that time. First, the neglect of situated language use (*parole*) was challenged by the inclusion of pragmatics and discourse or conversation analysis into linguistics. Spontaneous data began to be analysed in their own right, not just in terms of a deficient application of a rule apparatus. The beginnings of this pragmatic turn in linguistics were quite independent from research on bilingualism; they started in different branches of linguistics, heavily relied on other disciplines such as philosophy and sociology, and remained within a monolingual framework. However, they laid the grounds for viewing code-switching (a) as a performance phenomenon which nonetheless showed a good deal of regularity (Poplack 1980) and (b) as a discourse strategy which had its own discourse-functional meaning (Gumperz 1982). Second, socio-linguistics established itself as a new paradigm. Bilingualism became a social issue together with large-scale immigration movements into countries which so far had considered themselves to be monolingual, and with language rights movements insisting on the maintenance or even revitalization of 'old' minorities in the European nation-states. Third, generative linguists moved away from analysing idealized speaker/hearers and turned to the direct investigation of the principles of Universal Grammar; and since bilingual speakers doing code-switching/-mixing could be presumed to use such principles, bilingual talk provided access to UG (as do, for instance, creoles). The search for universal constraints of code-switching found its origin and theoretical justification in this line of thought.

All these changes, and the truly impressive amount of research they have triggered, have greatly increased our knowledge of bilingualism as a social and individual phenomenon. Nevertheless, I believe that there are at least three issues which continue to challenge linguistic thinking, all tied to the central question: can bilingual talk be reduced to two-sided monolingual talk? These issues, which are the topic of this chapter, are:

- (1) where does one language stop and the other start in bilingual talk?
- (2) where does code-switching/-mixing belong – to grammar or performance?
- and
- (3) are the languages used in bilingual talk the same languages that are used in monolingual talk?

Orthodox linguistic thinking as established over the history of linguistics presupposes a clear answer to the first and second questions, and a positive one to the third. It is only in this way that the assumptions about languages as bounded systems which have been foundational in the discipline can be upheld.

where does talk in one language ('code', variety) stop and talk in the other start?

Many investigations from the 1980s onwards (including Auer 1984; Li 1994) have shown that in certain speech communities, languages tend to be opposed to each other in systematically meaningful (i.e. discourse-functional) ways, and in such a way that neither the speakers nor the analysts have great doubts about assigning a given stretch of talk to one language or the other. Even in these communities though, there are certain stretches of talk which are of ambiguous or ambivalent status regarding their categorization as language A or language B. Studies on other bilingual communities, particularly those in which the two varieties in contact are structurally closely related, and those in which the use of the two languages is not discourse-functional (i.e. on code-mixing), have been added in recent years (cf., among many others, the contributions in Auer (ed.) 1998), and it has become more and more obvious that the clear-cut alternation between one language/variety and the other (code-switching in the strict sense of the term as used in Auer 1999) is but one case of bilingual talk.

Utterances may belong to both 'codes' in play and therefore be ambivalent. More precisely, the 'codes' may converge to such a degree that for the speakers and recipients of bilingual talk, no clear delimitation is possible.¹ Among the clear cases of ambivalence are so-called homophonous diamorphs or their approximations, but also discourse markers such as hesitation particles (*uhm*) or agreement markers (*mm*), and proper names.² Since Michael Clyne's 1967 book, we know ambivalent stretches of talk may facilitate or even trigger switching from one language into the other. Examples (1) and (2), below, show two cases of such a triggering (or, in a less deterministic way of speaking, facilitating) of code-switching by near-homophonous diamorphs.

- (1) Clyne (1980: 400): German descent immigrant in Australia

wir nehmen unsre biecher fär vier (-) for four periods
we take our books for four
'we take our books for four – for four periods'

Here, the near-homophonous diamorphs are dialectal German /fer/ and Australian English /for/ as well as German /fi:r/ and Australian English /for/.³ The closeness of German and English facilitates switching into English, the

dominant language of this speaker, which in itself is likely to be due to the non-availability of the German equivalent of English *periods*. In the transcription, italics indicate what would count as English according to the linguist's judgement. This follows common usage in bilingualism research which enforces a dichotomic opposition between the two languages. But triggering shows that the perceived distance between the two languages is not always the same (as it should be the case between two language systems), but rather a matter of degree.

- (2) Bierbach and Birken-Silverman (personal communication, 2003): male Italian guestworker in Germany, old-age pensioner in Germany
- what eat what drink all chemistry
'what we eat, what we drink, it's all chemical, all chemical, there is nothing natural'
was essen, was trinken, **alles chemie** (-) **non c'è niente naturale**
NEG there is nothing natural
- Here, the near-homophonous diamorphs are German /kemi:/ and Italian /kimika/. The dominant language of this speaker is Italian (his German exemplifies what has been called a 'guestworkers' pidgin'), and he uses the closeness of Italian and German in order to switch into the latter. The transition is smoothed by the repetition of *alles chemie* in a mixed version (German quantifier, Italian head in the noun phrase *alles chimica*). While these examples seem simple enough (and abundant in the literature; cf. Clyne 1987; Muytsken 2000), they already indicate an important problem: it is not the linguists who decide on how convergent two structures need to be in order to facilitate a bilingual speaker's switch from one language into the other, but rather that the decision is up to the bilinguals themselves. It thereby turns out to be a matter not only of structural distance but also of the perception of this distance between the 'codes' which in turn may be subject to attitudinal matters, including language ideologies.
- There is an additional issue. In the two examples cited above, the near-homophonous diamorphs do not differ from the respective words in monolingual talk. However, bilinguals do not usually speak like two-fold monolinguals. Rather, the 'codes' they oppose in 'code-switching' are THEIR variants of the respective monolingual codes. In the discussion of the two examples, we assumed that the convergence is purely linear, or syntagmatic (*in praesentia*), and that it is this on-the-spot convergence which creates fuzziness in the transitions from one language/'code'/variety into the other (and maybe even triggers it). This, however, is not necessarily the case, since the systems themselves can converge (paradigmatically, or *in absentia*). The simplest case is lexical convergence: bilinguals copy lexical items from language B into language A. Consider the following example:

(3) Bierbach and Birken-Silverman (personal communication, 2003): female middle-aged Italian guestworker born in Sicily, living in Germany

basta ca c'è frau diener
enough that there-is Ms Diener
si rump tutt- u scist.
REFL breaks whole the shift

die macht die ganz atmosphäre da drin
she makes the whole atmosphere there in-there

'if there wasn't Frau Diener the whole shift would break down. She makes the whole atmosphere in there'

While the third turn component of this utterance is clearly dialectalized German, and *Diener* a German family name (and as such outside of what can be translated), the middle turn component contains an element which could be classified either as an integrated loan or as an ad hoc lexical insertion from German into Italian (dialect): *scist 'shift'*. Accordingly, using italics for German and normal font for Italian, the utterance might just as well be transcribed as:

si rump tutt-u *scist*.
REFL breaks whole the shift

Scist derives from the German word *Schicht* /ʃɪçt/ in its local dialectal form with a coronalized fricative instead of the standard palatal one: /ʃɪf/. Syllable-final /ʃ/ is not well-formed in Sicilian and the palatal fricative therefore turns into an alveolar. This phonological adaptation ('integration') might be taken as an argument for *scist* having been copied into the variety of Sicilian/Italian spoken in Germany by immigrants/guest workers. As part of this variety, it may 'trigger' the switch into German, since it is a homophonous diamorph of German *Schicht*. However it is well known that phonological integration also occurs ad hoc and is not restricted to established loan-words, and that established loans may surface in their quasi-original shape (see example 6 below and the subsequent discussion). If we consider *scist* to be an ad hoc borrowing, then we are dealing not with convergence between two distinct language systems, but with an example of insertional mixing preparing alternative code-switching (cf. Auer 1999 for this terminology).

Here is another example:

(4) bilingual Italian adolescents in Germany; data from 1980s; Auer (1983: 337)

quanda sono sceso de quella scuola da a **hauptschule**;
when to-be-1.Sg. descend-PART from this school from the 'secondary modern' (= the lowest level in the German school system)
'when I left this school the "secondary modern"'

sane venuto in quella scuola scola
to-be-1.Sg. come-PARTICIPLE in this this school
'I came into this other school' (= the more prestigious, middle level in the German school system)

u deutsch a mi (-) era uguale com=enda a **hauptschule**;
the German for me – was same as in the 'secondary modern'
German (as a subject) was the same for me as in 'secondary modern'

gar keine (prob) (-)
PARTICLE no (problem)
'no problem at all'

wir schreiben keine dikte nix;
we write-1.PLPräs. no dictations nothing
'we write no dictations, nothing'

Again, there is a code-switch from Italian (dialect) into German. Two words in the Italian passage preceding the switch may have facilitated this transition: the word *hauptschule* (a particular German school type), and the word *deutsch* as a school subject. In both cases, the fact that there are no equivalents in Italian supports their classification as part of these Italian adolescents' variety of Italian. On the other hand, they may also be analysed as ad hoc borrowings, i.e. as a case of code-switching (or -mixing) themselves. Under one interpretation, the convergence is paradigmatic, under the other, it is syntagmatic. Issues of indeterminacy of this kind are a notorious problem for studies on code-switching/-mixing, rendering the idea that bilingual talk can (as the term 'bilingual' suggests) be studied by separating stretches of talk in one language from those in the other problematic. I will argue in the next section that there is no way of drawing a clear boundary between integrated borrowings on the one hand (which could then function as diamorphs for their 'monolingual' counterparts in the source language, and hence serve as facilitators of language alternation), and language switching of a single lexical item (which of course cannot be its own facilitator) on the other. In the section after that I will show that grammar is also affected. Here, the point is that there is no monolingual 'area of security' to which we linguists can turn in order to describe and make sense of the bilingual data: the monolingual speakers and their grammars which we tend to refer to are of no help.

Various researchers on bilingualism have shown that ambiguous stretches of talk between language A and language B not only 'happen' but are *made* to happen by the speakers. Speakers may, for interactional reasons, wish to produce a conversational contribution which cannot be classified as language A or language B. The most straightforward way of doing so is the intra-turn translation of an utterance into the other language, with the effect of performing (more or less) the (same) action twice, but in two languages. Monica Heller (1988) has described code-switching as a way to achieve 'strategic ambiguity' (and neutrality) in certain contexts in such a way. For structural linguistics, these cases are unproblematic since despite the fact that the speaker wants to keep language choice open, the turn-internal stretches of talk in language A and language B are easy to identify. More difficult are grammatical units in which elements from two different languages are conjoined by ambivalent elements which can be part of both. Kathryn Woolard (1987, again in 1999: 7) gives an example of an intentionally hybrid utterance of this kind. It is the standard opening phrase of a Catalan comedian (Eugenio):

(5) Woolard (1999: 7)

el **saben** aquel...
he know-3.PL. this-one...
'do you know this one?'

It draws its humour exactly from this hybirdity: the verb *saben* is either Catalan or Castilian, while the third person pronoun *el* is Catalan and the object pronoun *aquel* is Castilian. Between these elements the neutral *saben* functions as a pivot, making the utterance as a whole impossible to classify as either Castilian or Catalan, and the transition from Catalan into Castilian impossible to locate exactly.

In the examples discussed so far, the ambiguous stretches of talk were short. However, the same ambivalence may be found over larger units as well. Celso Alvarez-Cáccamo (1990) reports on Galician/Castilian bilinguals in Galicia who use Galician prosody with Castilian lexical items and syntax in certain situations. Since the two varieties are very close in structural terms, prosody is enough to make Castilian grammar and words sound Galician. The speaker thereby camouflages their use of Castilian without wholeheartedly speaking Galician. The structural hybridity of talk is mirrored by its social-interactional neutrality.

borrowings: grammar or discourse?

I have started out with a very simple observation, i.e. the fact that the exhaustive, non-residual segmentation of bilingual talk into utterances in one language and utterances in the other is not always possible, or at least

more problematic and more presupposing than is often thought. I have used the structuralist terms 'paradigmatic' and 'syntagmatic' to refer to two kinds of convergence which stand in the way of such a segmentation. It is tempting to conclude that paradigmatic convergence is tantamount to convergence of two systems, while syntagmatic convergence is a discourse phenomenon. This, however, is not true. There is a continuum of insertions with ad hoc (nonce) borrowing at one extreme, and sedimented borrowings (words which are habitually used by a certain speaker or even in a bilingual community) at the other extreme. Only in the second case is the borrowing part of the 'system' of the receiving variety (i.e. shared knowledge in the speech community).

While the extreme points of the continuum are relatively easy to describe, there are many intermediate positions on this scale which are less obvious. Many researchers on bilingualism therefore argue that a clear boundary between discourse (nonce borrowing) and the language system (lexical loans) cannot be defined (cf. Treffers-Daller 1994; Gardner-Chloros and Edwards 2004; Backus 1996; Angermeyer 2005: 325). The most obvious criterion, i.e. phonological and morpho-syntactic integration into the receiving language, is not only gradual in nature itself, it also does not yield satisfactory results: as mentioned before, ad hoc insertions can be integrated just like established loans. Take the following example of a French verbal stem (*cueill-*) inserted into an Alsatian (dialectal) sentence:

(6) Gardner-Chloros and Edwards (2004): French-Alsatian code-switching
in Strasbourg

ah voilà, mitt dass se do **cueillier**, un gehen dann uf d'ander Sit
PART not that they there pick-INF, and go then to the other side
'yes there you are, they shouldn't pick, and then go to the other side'

There is no established loan word *cueillier(e)* in Alsatian. The word is made up from a French verbal stem (cf. French *cueill+ir*) and a dialectal infinitive ending (-*ier* = /i:r/) and inserted into an Alsatian sentence frame. In this frame, a non-integrated French verb would sound awkward. In order to make the insertion well-formed, the speaker uses a German (and Alsatian) strategy which dates back hundreds of years as a way of integrating Latin and French stems, i.e. the suffix -*iere(n)*.⁴ This strategy itself is a French borrowing, but established in German and integrated into its morphology since the Early New High German period. The result is the ad hoc production of a compromise form in which there is a mixture between the two languages: a verb with a French stem and a Germanic suffix.

It may be objected that morphological integration was facilitated by the strong morphological similarity of the two languages in this particular case (the infinitives -*ir* and -*ter*, both /i:r/). This surely cannot be said of the following example of an English ad hoc insertion into a Russian frame:

(7) Angermeyer (2005:329)

čto vy duma-ete nepravil'no s ètim estimate-on?
 what you think-2Pl. wrong with this-INSTR estimate-INSTR
 'what do you think is wrong with this estimate?'

Again, *estimate* is not an established loan in Russian. However, the speaker makes the noun fit into the Russian sentence frame by adding the instrumental case to it, as it ought to be with a Russian noun.

The matter of integration has been discussed here in a very simplistic way, referring to surface integration by morphological means only. Of course there are more subtle methods of integration, both in phonology and phonetics and in grammar. With regards to the latter, Poplack and colleagues (e.g. Poplack and Meechan 1995) have shown that insertions which are not surface-marked for integration into the matrix language can nonetheless be syntactically integrated by obeying the structural requirements of word order of the surrounding language. In all these cases of integration of an ad hoc item, the item is used according to the system of the surrounding language without actually becoming a part of it (in this case, a part of the vocabulary).

But what does it mean to say that an insertion is or is not part of the linguistic system of the language into which it is inserted? If structural adaptation does not provide a feasible method to establish the loan status of an insertion, what we are left with is frequency of use. However, it is obvious that frequency of usage is nothing but a very superficial generalization over individual usage patterns. While some speakers may use the inserted items regularly, others may not even know it. While some speakers may use the inserted item to replace an item in the receiving language, others may give it a specific meaning, etc. It certainly does not provide an explanation of how items are copied from one language and integrated into the other. To find such an explanation, the underlying steps (practices) which lead to social diffusion have to be investigated.

I want to suggest that at least in the beginning of the structural sedimentation of a lexical insertion from language B into the system (lexicon) of language A, discursive strategies can be identified. Regularly, there is a difference between the first use of an (ad hoc) borrowing within a conversation, and the subsequent uses of the same word. That is, the first insertion is marked as an unexpected language choice, and thus metapragmatically framed as a deviation from the 'proper' (i.e. cooperatively established) language choice at that moment. Later the same word is used without such marking. For instance, in the following extract from a group discussion with Italian adolescents in Germany, one of the boys uses for the first time in this interactional episode the German word *Lehre* in an Italian (dialectal) context. (The Italian equivalent for 'apprenticeship' would be *appendistato*. We do not know whether it was known to the speaker – chances are it was not. However, the

word was known in the community and used by (some) adult Italians. This is reflected in the way the German insertion is treated in a conversation in which an adult Italian was present.)

(8) Italian/German/Italian dialect

- Again, *estimate* is not an established loan in Russian. However, the speaker makes the noun fit into the Russian sentence frame by adding the instrumental case to it, as it ought to be with a Russian noun.
- The matter of integration has been discussed here in a very simplistic way, referring to surface integration by morphological means only. Of course there are more subtle methods of integration, both in phonology and phonetics and in grammar. With regards to the latter, Poplack and colleagues (e.g. Poplack and Meechan 1995) have shown that insertions which are not surface-marked for integration into the matrix language can nonetheless be syntactically integrated by obeying the structural requirements of word order of the surrounding language. In all these cases of integration of an ad hoc item, the item is used according to the system of the surrounding language without actually becoming a part of it (in this case, a part of the vocabulary).
- But what does it mean to say that an insertion is or is not part of the linguistic system of the language into which it is inserted? If structural adaptation does not provide a feasible method to establish the loan status of an insertion, what we are left with is frequency of use. However, it is obvious that frequency of usage is nothing but a very superficial generalization over individual usage patterns. While some speakers may use the inserted items regularly, others may not even know it. While some speakers may use the inserted item to replace an item in the receiving language, others may give it a specific meaning, etc. It certainly does not provide an explanation of how items are copied from one language and integrated into the other. To find such an explanation, the underlying steps (practices) which lead to social diffusion have to be investigated.

I want to suggest that at least in the beginning of the structural sedimentation of a lexical insertion from language B into the system (lexicon) of language A, discursive strategies can be identified. Regularly, there is a difference between the first use of an (ad hoc) borrowing within a conversation, and the subsequent uses of the same word. That is, the first insertion is marked as an unexpected language choice, and thus metapragmatically framed as a deviation from the 'proper' (i.e. cooperatively established) language choice at that moment. Later the same word is used without such marking. For instance, in the following extract from a group discussion with Italian adolescents in Germany, one of the boys uses for the first time in this interactional episode the German word *Lehre* in an Italian (dialectal) context. (The Italian equivalent for 'apprenticeship' would be *appendistato*. We do not know whether it was known to the speaker – chances are it was not. However, the

word was known in the community and used by (some) adult Italians. This is reflected in the way the German insertion is treated in a conversation in which an adult Italian was present.)

(9) subsequent uses of *Lehre* by the same speaker, in chronological order

- (a) che c'ho a practica roppa è più meglio
 that I-have the practice then is more better
 'I thought that perhaps first I'll go on a (-) an apprenticeship, so that I have more practice, this is better'
- (b) che= hai fatt= a Lehr(.) continua a scuola
 after that-CON have-2Sg made-PART the apprenticeship continues the school
 'after you've finished your apprenticeship, school continues'
- (c) prima=a Lehre andá u tege
 first an apprenticeship go-INF the T.G.₅
 'first an apprenticeship – then grammar school'
- (d) fare prima na Lehr; (-) qu= è più meglio e que prendo a praxis;
 do first an apprenticeship that-CON It is more better and that-CON take-2.Sg. the practice
 'do an apprenticeship first; so that it is better and so that you have practice'
- (e) più meglio a fa prima na Lehr che ro ciòé veramend=a praxis;
 more better to do first an apprenticeship so that then PART really the practice
 'it is better to do an apprenticeship first so that then (you're) really (got) practice'

Here, we can observe sedimentation *in statu nascondi*. All uses of *Lehre* are insertions into an Italian dialectal matrix, of course; but the more the speaker

repeats the word, the less it is treated as an 'intruder' into Italian, and the less its usage becomes problematic from the point of view of language choice. It goes without saying that a fully established loan word would not be framed in such a way.

In a recent dissertation on bilingual talk in New York small claim courts involving translators, Angermeyer (2005; also cf. Angermeyer 2002) has shown that other-language insertions of this kind can be used in order to create cohesion. He observed that the vast majority of lexical insertions used by the interpreters are words that had been used by one of the participants before, usually in the source text to be translated. Sometimes it also was the litigant or claimant who inserted items used by the translator in his or her subsequent turns, as in the following example:

(10a) Angermeyer (2005: 28), simplified; claimant wants be compensated by landlord for having had to hire an exterminator for her apartment

- | | | |
|-----------|--|---|
| Claimant: | Es es de ((shows a receipt))
‘that’s for...’ | ↗ |
| Interpr.: | es de
‘it’s for...’ | |
| Claimant: | eso es del del fumigador [que fue
‘that’s for, for the exterminator who came’] | |
| Interpr.: | <i>who [went to the apartment
[exterminator
ah, sorry
exterminator
Interpr.: [desculpe
‘sorry’</i> | |

There is no reason to believe that *exterminator* is an established loan in Spanish as spoken in NYC (in fact, the claimant uses the word *fumigador*). However, once the translator has introduced the English term (in 10a), the claimant uses this term herself as an insertion in her Spanish utterances; see the following extract from the same proceedings, which occurred after (10a):

- | | |
|---|---|
| (10b) Angermeyer (2005: 28), same participants as (10a) | |
| Arbitrator: | <i>what else are you [s</i> |
| Interpr.: | <i>[que yo tuve que
‘that I had to’</i> |
| Arbitrator: | <i>sung [for besides this and this?</i> |
| Claimant: | <i>[eh que que
‘uhm to-to’</i> |

As in the Italian example, the first occurrence is marked (here in a dialogical fashion: the Spanish-speaking claimant takes up the interpreter's translation of his *fumigador* into English by repeating it twice, as if she was eager to memorize it), while in the subsequent uses the newly acquired or remembered word is used as if it had always been part of the speaker's repertoire. Examples such as these mark the beginning of 'integration' – a discursive strategy. Individual use is the starting point, and a loan is nothing more than the consequence of the repeated use of an item.

I have argued for a continuum of lexical insertions from ad hoc (nonce) borrowings to established loans. But if this is true, the question must be asked: where does switching/mixing belong – to the language system or to discourse? The question links up with the old dilemma about the nature of grammatical constraints on bilingual talk. Many researchers have pointed out that code-mixing is subject to certain (language specific) structural constraints, and have described structured ways of making the two grammars compatible. These clearly reflect some kind of linguistic knowledge (competence) on the part of the speakers. On the other hand, only rarely have linguists written 'grammars' of code-mixing (in a non-metaphorical sense of the word), since (apart from extreme cases such as fused languages like Michif or Copper Island Aleut; cf. Bakker 1994) mixing, although constrained, is not obligatory and predictable, and therefore outside the reach of grammar (in the sense of the *ars obligatoria*). The problem is not all that unfamiliar to linguists who work on grammaticalization, or in some version of exemplar theory in phonology, both of which question the clear division between grammar and discourse. So here we have a second area of indeterminacy.

the two codes of code-switching/-mixing

The examples of paradigmatic convergence discussed above concerned lexical insertion. Most researchers on bilingualism agree that insertion requires, by its very nature and definition, a grammatical frame of the receiving language into which the borrowed item is inserted. It is generally concluded that although a high number of insertions of language B elements may occur in a grammatical unit (such as a sentence, clause or complementizer phrase), they cannot change its belonging to language A. There remains a neat imbalance between what Myers-Scotton (e.g. 2002) calls the 'matrix language' which provides the grammar of the unit as a whole, and the inserted elements which find

their place in this matrix. Grammar (morphology and syntax) tells us which language a grammatical unit 'really' belongs to, the lexicon is superficial. This is a conviction which is as old as linguistic research on language contact. As early as 1898 Hermann Paul expressed it as follows:

Er [der Einzelne] wird vielleicht, wenn er beide [Sprachen] gleich gut beherrscht, sehr leicht aus der einen in die andere übergehen, aber innerhalb eines Satzgefüges wird doch immer die eine die eigentliche Grundlage bilden, die andere wird, wenn sie auch mehr oder weniger modifizierend einwirkt, nur eine sekundäre Rolle spielen. (1898: 366)
 ('He [the individual] who masters the two [languages] equally well, will perhaps very easily go from one into other, but within a sentence construction, there will always be one [language] which forms the real basis, the other will play a secondary role, even though it may be more or less a modifier.'

More than a hundred years later, Myers-Scotton echoes this conviction when she writes:

One of the languages involved in C[ode]S[witching] plays a dominant role. This language is labelled the Matrix Language (ML), and its grammar sets the morphosyntactic frame [...]. (1993: 229)

Let us therefore look at grammar, and pursue Myers-Scotton's argument somewhat further here.

The matrix language in this model provides (most of) the grammatical morphemes and the word/morpheme order of the clause as a whole. Myers-Scotton's theory acknowledges that some morphological elements from the embedded language may nonetheless appear in a sentence (so called early system and bridge system morphemes). What must be provided by the matrix language, however, are system morphemes which have grammatical relations external to their head constituent (e.g. those which assign or receive thematic roles) such as morphemes taking part in subject/predicate congruence, case, tense and aspect morphology (cf. Myers-Scotton 2002: ch. 3).

There are, however, clear counter-examples to this claim. In this case, a constellation of morphemes (including late system morphemes) is taken over as a whole and inserted into the matrix languages. Myers-Scotton calls these complex insertions 'embedded language islands' and postulates that their internal make-up follows the grammar of the embedded language. Embedded language islands occur, among other things, when the grammar of the matrix language and that of the embedded language are non-congruent. She gives the following example:

(11) Myers-Scotton (2002: 99): English/Swahili, data from 1988

a-na-i-tumi-a
 3sg-nom-past-obj-'use'-FV
 'he uses it for personal purposes' (refers to his car mentioned in previous discourse)

Here, the set prepositional phrase *for personal purposes* is an island within a Swahili sentence. The island occupies the position of an adverbial phrase in Swahili, but its internal syntax is well-formed according to English syntax, not according to Swahili syntax.

The model is attractive because it predicts a neat separation of the domains of the two grammars in contact. One is responsible for the matrix, the other for the insertions, even when they are internally complex. However, Raihan Muhamedova (2006) has described code-mixing between Kazakh and Russian and between Uighur and Russian in Almaty which systematically contradicts this pattern. Among the systematic structures she identified which have a direct bearing on the question of embedded language islands we find nominal Russian islands (NPs) in Kazakh (or Uighur) complementizer phrases. Cf. the following extract:

(12) Muhamedova (2005: 83–4): Kazakh/Russian⁶

Men me () meniň bala-m bar, žetpis alpis-nši žil-ği.
 1 1 my child-POSIP EXIST seventy-six-ORD year-ADJ
 1 (.) have a child, born in seventy six

өүсүү-үү **школ-ди** bitur-d osında
 high ?? school-ACK finish-3PS here
 He finished high school here

(.) *мранчнорм-түү* milicija⁷-da iste-di.
 traffic-SUFF-NOM.Sg.MASK police-LOK work-3Ps.
 and then he worked here with the traffic police

The speaker inserts the complex Russian noun phrase высш-ий школ into a Kazakh matrix frame, which receives the Kazakh accusative suffix *-di*. However, in contradiction to the prediction by Myers-Scotton's model, this embedded language island is not well-formed in the donor language, Russian: in monolingual Russian the noun phrase would be высш-ая школа. Two differences between monolingual and 'bilingual' Russian are noteworthy. First, in the insertion the Russian head of the NP, школа, has lost the final vowel (truncation). This is remarkable since the suffix carries the most important morphological information in Russian (such as case, gender and number in the noun). Thus, while the Russian suffix /-a/ marks the word as a feminine noun, its truncated equivalent as used in code-mixing cannot

be assigned gender any longer (or, if the rules of monolingual Russian are applied, its gender is now masculine). Second: the adjective modifying the head carries a suffix, but from the point of view of monolingual Russian, it is the masculine suffix (*/ij/*) instead of the feminine one (*/aja/*) which is required to achieve congruence with the (monolingual Russian) feminine noun школа. What has happened? Saying that the noun's gender has changed from feminine to masculine would be monolingually biased. More plausibly, the Russian feminine noun has lost its gender entirely, and the /ij/-suffix in высту-ий has become the unmarked form of adjectival ending. At this point it is important to know that Kazakh has no gender. This means that, as the matrix language, Kazakh not only governs the external syntax of the insertion (i.e. its position within the sentence and its case, cf. the suffix), but it also influences its internal structure: the Russian used in the embedded language island has converged with Kazakh. It is therefore substantially different both from monolingual Russian as spoken in Russia, and also from Russian spoken by the same informants in Almaty who only rarely speak simplified Russian (without gender) in monolingual talk.⁸

This is not a singular case. Take the following, somewhat more complex example from the same data set:

(13) Muhamedova (2005: 85): Kazakh/Russian

Sosin ne (–) o oris заетәйиң-иүү-лар ke-p
 then uhm (–) r Russian director-NOM.SG.MASK-PI come-CONV
 then uhm (–) when the directors (of the library) come

žaňagi книж-н-бүү өйткәв-к-лар tema-lar-i-n ne-ler-i-n
 this/uhm book-ADJ-NOM.SG.MASK exposition-?Sg.FEM-PI theme-PI.POSS3P-AKK thing-PI-
 POSS3P-AKK

sura-ğan-da, ...
 ask-PA RT-LOK
 and uhm ask for the themes and things of the book expositions ...

Again, the matrix language of this CP is Kazakh. The complex insertion we are interested in is the noun phrase made up of a modifying adjective and a head noun: книж-н-ый выставк-а. Once more, the Russian island (or rather, its head noun?) additionally receives a Kazakh suffix (the plural *-lar*). In monolingual Russian, 'book expositions' is книж-н-ые выставк-и, with the plural obligatorily marked on both the head and the modifier (the singular is книж-н-ая выставк-а, feminine marked both on the head and the modifier). However, the island produced by the bilingual speaker has the structure shown in Figure 15.1; or alternatively, in Figure 15.2. (The thick bracket marks non-congruence from the monolingual Russian point of view.)

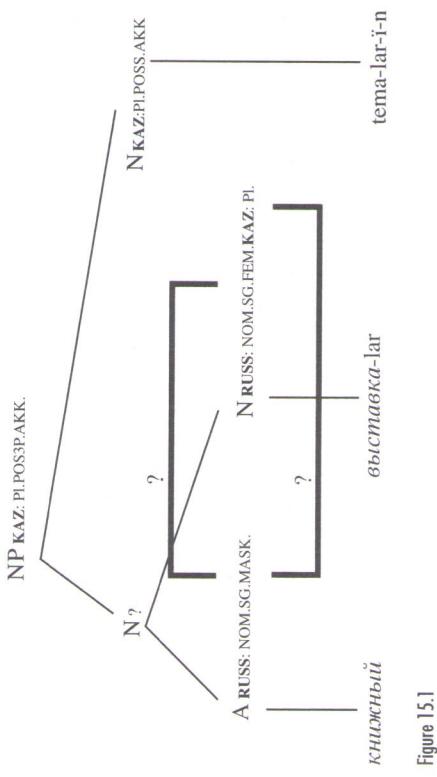


Figure 15.1

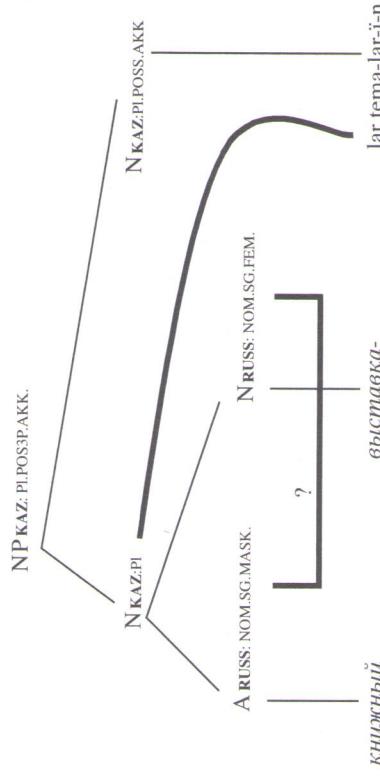


Figure 15.2

Under both analyses, the NP книж-н-ый выставк-а fails to satisfy the requirement that islands be well-formed in the embedded language. As in the first example, the adjective receives a suffix which, according to Russian grammar, is masculine, while the head noun has a feminine ending. Again, there are good arguments that the latter is not analysed as a gender suffix any longer, and that the former is a gender-neutral adjectival suffix. In addition, example (13) shows a change in the number-marking system of Russian. While monolingual Russian requires number to be marked both on the head and on the adjectival modifier, Kazakh marks plural on the head only. книжный выставка-*лар* is a plural NP, which is marked according to the Kazakh system on the head noun (using the Kazakh suffix).⁹ The modifier is unmarked for number. In sum, all the morphology of this island is Russian, yet the way in which this morphology used is not Russian, but Kazakh. Is the embedded NP then Russian or Kazakh?

It turns out that the neat separation between the two languages which the distinction between matrix and embedded language seems to guarantee is not achieved in Myers-Scotton's model. The author gives another example for this by pointing out that numerous cases are attested in which the matrix language fails to provide the morphology of a single inserted lexical item, i.e. in which the latter is strong enough to 'ward off' matrix language integration. As a consequence, 'bare forms' may be inserted into a matrix language frame which would require morphological marking. They are ungrammatical both in the matrix language and in the embedded language. Many examples for this can be found in Louis Boumans's study on Moroccan Arabic/Dutch code-mixing, for example:

(14) Boumans (1998: 218)

mša *discothek*
go disco
'we went to the disco'

Arabic as the matrix language would require a preposition (*/l-*) and a definite prefix on the noun; however, the speaker integrates Dutch *discothek* without such a marking. Myers-Scotton (2002: 66) therefore insists on not equating the matrix language with any existing language, let alone the monolingual contact language (*ibid.*) – it is just an 'abstract frame for the morphosyntax of the bilingual CP'. While bare forms question the identity of the matrix language with the respective monolingual language, the Kazakh/Russian examples given above question the identity of the embedded language with the respective monolingual language. Here, too, we need to insist on the separation of the embedded language and the monolingual contact language. This has far-reaching consequences. For instance, it is unclear how notions such as the Myers-Scotton's 'congruence' between the matrix and the embedded language (the lack of which is made responsible for the occurrence of islands) should be defined. It is even unclear how morpheme order and late system morphology can be established without reference to the corresponding monolingual variety; this is crucial, since of the identification of the matrix language relies on it: if the latter cannot be equated with the monolingual code, we have no means of identifying it. The conclusion, however, is inevitable: bilingual talk cannot be analysed as a mixture of two monolingual codes. An alternative has been formulated in a 1998 paper by Mark Sebba in which he shows that congruence is indeed a condition for mixing, but that this congruence is not defined by the linguist looking at monolingual codes, but by the bilinguals themselves, who achieve it by strategies such as 'harmonization', 'neutralization', 'compromising' and 'blocking': 'Bilinguals "create" congruent categories by finding common ground between the languages concerned' (Sebba 1998: 8). Sebba argues that a bilingual community establishes the congruence of the categories of the languages involved the longer it exists; or, in other words,

for this by pointing out that numerous cases are attested in which the matrix language fails to provide the morphology of a single inserted lexical item, i.e. in which the latter is strong enough to 'ward off' matrix language integration. As a consequence, 'bare forms' may be inserted into a matrix language frame which would require morphological marking. They are ungrammatical both in the matrix language and in the embedded language. Many examples for this can be found in Louis Boumans's study on Moroccan Arabic/Dutch code-mixing, for example:

conclusion

In this chapter I have argued that the traditional approach to bilingual talk as the combination of two pre-existing language systems is hard to reconcile with syntagmatic and paradigmatic cases of 'code convergence' and therefore leads to a number of difficulties. In the second section, syntagmatic and paradigmatic 'code convergence' were shown to result in factual and/or perceived stretches of ambivalent talk between language A and language B. In the next section, I have argued that paradigmatic convergence in the lexicon (borrowing) is best analysed as an ongoing process of sedimentation based on ad hoc insertions the beginnings of which can be observed as interactional strategies within an interactional episode. The section after that showed that the distinction between matrix language and embedded language is useful but that the two must not be equated with the respective monolingual varieties since structural convergence may take place between them which does not occur in monolingual speech (even by bilingual speakers). As a consequence, I suggest that the starting point of bilingual analyses can no longer be two languages, but rather a collection of discursive and linguistic practices used by bilingual speakers in a community, and based on certain grammatical/lexical/phonological feature constellations. These constellations may be opposed to each other for functional reasons (as in discourse-functional code-switching), or they may stand side by side and constitute one speaking style (and therefore one 'code'), or they may become amalgamated into a new system such as in the so-called 'mixed' (or fused) languages (cf. Auer 1999). Seen from this perspective, the separation of two 'codes' in terms of historical languages becomes a secondary issue.

The answers to the questions that I raised at the beginning of this chapter – i.e. (1) where does one language stop and the other start in bilingual talk? (2) where does code-switching/mixing belong, to grammar or performance? and (3) are the languages used in bilingual talk the same languages that are used in monolingual talk? – are therefore hard to give in the first two cases, and negative in the third. With respect to (1), it seems that the bilingual speakers can choose to make the issue relevant or irrelevant, and that in both cases, ambiguous stretches of talk may result. With respect to (2), there is evidence that sedimented patterns of mixing (including what is known as borrowing) have their origin in discursive practices which may or may not make their way into the language 'system' (in terms of the shared linguistic knowledge in a community). Together with the negative answer to the third question, these results suggest that the assumption of bound linguistic systems as the object of linguistic research is questioned by bilingual practices. This is the challenge which bilingualism continues to represent for linguistics.

notes

1. Hill and Hill (1986) use the structuralist term of syncretism (irrelevance of an opposition) to refer to such linguistic objects.
2. With the exception of royalty and popes, a name designating a certain person does not translate into its equivalent (if there is one) in the other language; an Englishman by the name of *Charles* cannot be called in German by the equivalent of this name in this language, *Karl*; the translation would destroy the referential force of the proper name, and the speaker would not be heard as referring to 'Charles' any longer.
3. Depending on the phonetic realization of the words, particularly the final /r/, the objective contrast between the languages may be more or less pronounced.
4. The dialectal form has deleted the standard German final -n (as in all unstressed -en#-sequences) and in the present case, presumably due to an allegro rule, the -e preceding this -n as well. Note that the word is thereby made to sound very much like the French equivalent (from which it is only distinguished by the quality of the /r/).
5. 'Technical Gymnasium', a form of secondary school.
6. While Kazakh is usually written in Cyrillic, it is transliterated here, partly to highlight the difference between the two languages, and partly because Cyrillic is not well-suited for Kazakh phonology.
7. *Miftija* is an established Russian loan word in Kazakh.
8. As Muhamedova shows, the restructuring of Russian as a genderless language is quantitatively the dominant pattern in Kazakh/Russian code-mixing, although it rarely occurs in monolingual Russian spoken by the same informants (also see Auer and Muhamedova 2005).
9. Note that the Kazakh number suffix is not attached to the Russian stem, but to the Russian word form in the feminine singular. The stem is reanalysed from выстравка as выстравка.

references

- Alvarez Cáccamo, Celso (1990). Rethinking conversational code-switching: Codes, speech varieties, and contextualization. In *Proceeding of 6th Annual Meeting Berkeley Linguistic Society. General Session and Parasession on the Legacy of Grice*. Berkeley: Berkeley Linguistic Society, 3–16.
- Angermannayer, Philipp (2002). Lexical cohesion in multilingual conversation. *International Journal of Bilingualism* 6: 361–93.
- (2005). 'Speak English or what?' – Codeswitching and interpreter use in New York Small Claims Courts. Unpublished PhD thesis, New York University, Dept. of Linguistics.
- Auer, Peter (1983). *Zweisprachige Konversationen*. Unpublished PhD thesis, University of Constance.
- (1984). *Bilingual Conversation*. Amsterdam: Benjamins.
- (ed.) (1998). *Code-Switching in Interaction*. London: Routledge.
- (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3: 309–32.
- Auer, Peter, and Rathna Muhamedova (2005). 'Embedded language' and 'matrix language' in insertional language mixing: Some problematic cases. *Italian Journal of Linguistics* 17(1): 35–54.
- Backus, Ad (1996). *Two in One: Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: University Press.
- Bakker, Peter (ed.) (1994). *Mixed Languages: 15 Case Studies in Language Intertwining*. Amsterdam: IFONTT (Instituut voor functioneel onderzoek van taal en taalgebruik) at the University of Amsterdam.
- Boumans, Louis (1998). *The Syntax of Codeswitching: Analysing Moroccan Arabic/Dutch conversation*. Unpublished PhD thesis, University of Nijmegen.
- Clyne, Michael (1967). *Transference and Triggering*. The Hague: Mouton.
- (1980). Triggering and language processing. *Canadian Journal of Psychology* 34: 400–6.
- (1987). Constraints on code-switching: How universal are they? *Linguistics* 25: 739–64.
- Gardner-Chloros, Penelope, and Malcolm Edwards (2004). Assumptions behind grammatical approaches to code-switching: When the blueprint is a red herring. *Transactions of the Philological Society* 102(1): 103–29.
- Gumperz, John (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.
- Heller, Monica (1988). Strategic ambiguity: Codeswitching in the management of conflict. In M. Heller (ed.), *Codeswitching: Anthropological and Sociolinguistic Perspectives*, pp. 77–97. Berlin: Mouton.
- Hill, Jane, and Kenneth Hill (1986). *Speaking Mexican*. Tucson: University of Arizona Press.
- Le Page, Robert (1989). What is a language? *York Papers in Linguistics* 13: 9–24.
- Li, Wei (1994). *Three Generations, Two Languages, One Family*. Clevendon: Multilingual Matters.
- Muhamedowa, Raihan (2006). *Untersuchung zum Kasachisch-Russischen Code-Mixing (mit Ausblicken auf den uigurisch-russischen Sprachkontakt)*. Munich: Lincom Muyšken, Pieter (2000). *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol (1993). *Duellng Languages*. Oxford: Clarendon.
- (2002). *Contact Linguistics*. Oxford: Oxford University Press.
- Paul, Hermann (1998). *Prinzipien der Sprachgeschichte*. 3rd printing. Halle: Niemeyer.
- Poplack, Shana (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics* 18: 581–618.
- Poplack, Shana, and Marjorie Meechan (1995). Patterns of language mixture: Nominal structure in Welsh-French and Fongbe-French bilingual discourse. In L. Milroy and P. Muysken (eds), *One Speaker, Two Languages*, pp. 199–232. Cambridge: Cambridge University Press.
- Schuchardt, Hugo (1884). *Slawo-deutsches und Slavo-italienisches*. Graz: Leuschner & Lubensky.
- Sebba, Mark (1998). A congruence approach to the syntax of code-switching. *International Journal of Bilingualism* 2: 1–20.
- Treffers-Daller, Janine (1994). *Mixing Two Languages: French-Dutch Contact in a Comparative Perspective*. Berlin: Mouton.
- Weinreich, U. (1953). *Languages in Contact*. The Hague: Mouton.
- Woolard, Kathryn (1987). Codeswitching and comedy in Catalonia. *Pragmatics* 1: 106–22.
- (1999). Simultaneity and bivalence as strategies in bilingualism. *Journal of Linguistic Anthropology* 8(1): 3–29.