'Embedded language' and 'matrix language' in insertional language mixing: Some problematic cases

Peter Auer & Raihan Muhamedova

It is mostly assumed that the way in which languages may be combined within a syntactic unit is such that language A is dominant and language B (embedded language) is inserted (in the form of single words or of larger constituents) into the grammatical frame defined by language A (matrix language). The grammar of the matrix language provides the grammatical frame of the sentence as a whole, while the grammar of the embedded language is used only in complex insertions to determine the structure of the inserted constituent.

We will provide empirical evidence that, on the one hand, simple insertions do not necessarily take on the morphology of the matrix language, but may be treated according to the rules of the embedded language; and, on the other hand, that the language of complex insertions does not necessarily follow the grammar of the embedded language. The 'dominance' of one language over another in intrasentential code-mixing seems rather to be a matter of gradience and can mean different things in different situations of language contact.

1. Introduction

It is generally assumed (cf. Muysken 2000) that one way in which languages may be combined within a syntactic unit (e.g. a sentence or a clause) is such that language A is dominant and language B is inserted into the grammatical frame defined by language A. Following widespread use, we speak of *insertional* language *mixing* here (cf. Auer 1999a), and taking up a terminological proposal by Myers-Scotton (cf. 2002, among many other publications), we call the dominant language the *matrix* language and the inserted language the *embedded* language. The insertion itself may consist of a single lexical item, but it can also be more complex. In the latter case, many researchers (again following Myers-Scotton 1993; 2002) assume that the internal structure of the insertion follows the grammar of the embedded language instead of that of the matrix language. There is, then, a strict division of the realms of the two grammars: ideally, the grammar of the matrix language provides the grammatical frame of the sentence/clause as a whole, while the grammar of the embedded language is used in complex insertions.

In this article, we will show that such a position cannot be upheld. In particular, we will provide empirical evidence that (1) simple insertions do not necessarily take on the morphology of the matrix language,

but may be treated according to the rules of the embedded language which may even, in turn, influence the matrix language; and (2) that the language of complex insertions does not necessarily follow the grammar of the embedded language, at least if the latter is equated with a monolingual version of that language (as spoken by monolinguals of language A, or by bilinguals when they are speaking in a monolingual mode). As a consequence, we suggest that the 'dominance' of one language over another in intrasentential/intraclausal code-mixing is a matter of gradience and can mean very different things in different situations of language contact.

2. Insertions vs. Embedded Language Islands

The idea that two languages cannot mix in a balanced way in a bilingual clause [1] is an old one. As early as 1898, Hermann Paul declared (without giving much evidence for it):

> 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. (Paul 1898[3]: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.')

In the last decades, a number of researchers on bilingualism have put forward accounts of mixing which pursue the same line of thinking. Of them, Carol Myers-Scotton arguably has reformulated Paul's idea in the most sophisticated model so far. However, the basic idea has remained the same:

> One of the languages involved in CS [code-switching] plays a dominant role. This language is labelled the Matrix Language (ML), and its grammar sets the morphosyntactic frame [...]. (Myers-Scotton 1993:229)

We believe that this account is basically correct although we contend that there are other forms of code-switching and mixing beside the

insertional type (cf. Auer 1999a, Muysken 2000 for the distinction between insertion and alternation). These, however, are not the topic of this paper. The important (and controversial) point is of course to outline what it means that the grammar of the dominant language "sets the morphosyntactic frame." In Myers-Scotton's theory (in the latest version as of 2002), the matrix language provides the word/morpheme order of the clause as a whole as well as a certain class of grammatical morphemes, i.e. those which "have grammatical relations external to their head constituent" (Myers-Scotton 2002:87).[2] These are, for instance, morphemes taking part in subject/predicate congruence, as well as case, tense and aspect morphology. The following examples show cases of single lexical insertions which receive the late system morphemes of the matrix language, where necessary, and therefore are in accordance with this prediction:

(1)   (Türker 2000:70, Turkish/Norwegian in Oslo)
geç-en   sene   *serie*-de-ydi-k
pass-PART/SUBJ   year   league-LOC-bePAST-1PL
'last year we were in the league'

(2)   (Myers-Scotton 2002:89, Swahili/English in Nairobi, data from 1988)
ile m-geni,   hata si-ku-*comment*.
DEM/CL9 Ch/S-visitor,   even 1SG.NEG-PST.NEG-comment
'that visitor, I did not even comment'

In (1), Turkish is the matrix language and provides all the system morphemes, including the locative, tense and person/number suffixes on the predicate noun which is imported into the sentence from Norwegian. In (2), the matrix language is Swahili, not only because the sentence follows Swahili word order, but also because the system morphemes prefixed to the English verb are Swahili late system morphemes (person, number, tense). The inserted element is an English stem.

In cases such as (1) and (2), the matrix language exerts maximal control over the insertion. There are, however, cases where this control is less strong (all of which are acknowledged by Myers-Scotton as exceptions to the rule). First, there may be insertions which receive no marking at all, although the matrix language would require one in that particular slot. In example (3),

(3)   (Backus 1992:53, Turkish/Dutch in the Netherlands)
bu-n-lar   herkes kendi   *prijs*   söylü-yor
this-BUFFER-PL   everyone REFL   price   name-PROG.3SG
'here everybody names his own price'

the inserted Dutch noun occurs in a slot which, according to the grammar of the matrix language, Turkish, requires possessive and accusative marking. Both, however, are lacking. The dominance of the matrix language is not sufficient to overcome the resistance of the embedded language against Turkish inflection. Second, there are cases in which the inserted word carries along its own inflection. This may occur when the grammar of the matrix language and that of the embedded language are non-congruent. Late system morphemes may then combine with the inserted stem in order to form what Myers-Scotton calls an "embedded language island". Cf. the following examples:

(4)   (Stolt 1964:60, Latin/Early New High German)
       so   können   wir   nicht   *ecclesia-m*   lassen (5088b, Math.L.)
       so   can-1PL   we   not   church-AKK.SG leave-INF
       'so we cannot leave the church'

(5)   (Stolt 1964: 69, Latin/Early New High German; see below)
       omnes gentes, quae non habent religionem, mussen *superstitionem*   haben (371)
                                                   must-1PL superstition-AKK.SG. have-INF
       'all people who don't have a religion, must have a superstition'

Since the object case marking requires information from outside its maximal projection (i.e., from the nonfinite verb), the morphology of the object nouns *ecclesia* and *superstitio* should follow the rules of German as in *die ecclesia, eine superstitio*. Yet the Latin accusative suffix is used, forming a complex insertion. The rules of government of German are mapped onto the morphology of Latin: a German verb which governs a direct object requires the insertion from Latin in the object position to be marked by an accusative ending.

The notion of embedded language islands will play an important role in section 4 below. At this point, it should be added that their complexity can also be due to the fact that two or more content words are combined (with or without late system morphemes). Cf.:

(6)   (Stolt 1964:81, Latin/Early New High German)
       Human-um   *cor*   kann   es nit   fass-en (137)
       human-NEUTR.NOM/AKK.SG   heart-NOM/AKK   can-3SG it not   seize-INF
       'the human heart cannot grasp it'

The island here consists of a Latin subject noun phrase in which the adjective modifies the head (noun). The embedded language island is well-formed according to Latin syntax and morphology (and contains no German elements), and it occupies the position which it ought to accord-

ing to German syntax. Its internal syntax is entirely Latin, but its external syntax is entirely German.

(7)   (from: Boumans 1998:275, Moroccan Arabic/Dutch in the Netherlands)
op	*de*	*middelbar-e*	*school*	ʕend-hŭm	waḥed	*schema*
at	the	secondary-AGR	school	at-3PL	INDEF	scheme

'at the secondary school they have a scheme'

The general make-up of the sentence (without a finite verb of possession) follows the Arabic pattern, and most grammatical morphemes (such as the inflectional ending on the preposition and the indefinite pronoun) do so as well. *Scheme* is a single insertion into this matrix, but *op de middelbare school* is an internally complex noun phrase which is inserted sentence-initally into the Arabic clause. It includes at least one late system morpheme, i.e. the case suffx /e/, and is doubtlessly a well-formed prepositional phrase in monolingual Dutch.

(8)   (Treffers-Daller 1994:224, French-Dutch mixing in Brussels)
d'r	*zit*	*me*	*hier*	une	femme	qui	n'est	pas	drôle
there	sits	me	here	a	woman	who	NEG-is	not	funny

'here we have a woman who is not funny'

The subject noun phrase (an indefinite noun modified by a relative clause) is a French island within a basically Dutch matrix clause.

In all cases, the matrix language is not strong enough to dominate the whole sentence; it integrates utterance parts which are internally organised according to the principles of the embedded language. This is the division of labour as predicted by Myers-Scotton's model.

In the following sections, we will discuss cases beyond those which Myers-Scotton's model accommodates in order to show that the dominance of the matrix is gradient.

3. Latin insertions in Early New High German

In early modern times, many European intellectuals used a bilingual style for informal discussion and conversation in which elements of Latin (the dominant academic language) and the vernacular were mixed. The dinner table conversations of Martin Luther are a well-documented example of this style, since they were 'transcribed' by several of his co-present students. Although the available text records obviously are not *verbatim* transcripts, and differ according to the transcribers, they prove

that code-mixing occurred frequently. In addition, we can be sure that the structural regularities of code-mixing according to established use in humanist circles were obeyed in the transcriptions.[3] If anything, they may be biased in favour of a more written (and therefore more monolingual) style in which some cases of clause-internal language mixing were 'corrected'. What has survived editing represents well-formed cases of mixing for this group of speakers, even though they may not represent Luther's wording. In a ground-breaking but little known publication, Birgit Stolt (1964) studied these bilingual texts and gave the first structural account of code-mixing.[4] All the examples quoted here are taken from her book.

A first observation is that in modern terminology, Luther's dinner table speech is full of embedded language islands such as the ones given above (4-6). As Stolt argues, the status of Latin is so high that German as a matrix language cannot enforce its morphology on the insertions. However, Luther does not refrain from double-marking an inserted Latin noun both according to German and Latin. In this case, the noun receives congruent marking on the stem (Latin) and in the accompanying article (German), in the following example accusative singular feminine:

(9)   (Stolt 1964:60)
      Staupicius hat die   *doctrinam*   angefangen... (526)
      Staupicius has the-AKK.SG. doctrine-AKK.SG. start-PART
      'Staupicius started the doctrine'

This contradicts Myers-Scotton's model which only allows double marking in early system morphemes, not late system morphemes such as case. (Ex. (10) shows the same point.)

Of special interest are cases in which the mapping of the matrix language information onto the morphology of the embedded language is not straightforward. This is the case in some prepositional phrases with a German preposition and a Latin noun phrase. For instance, German can be mapped easily onto Latin in

(10a) (Stolt 1964:140)
      aus   den   *doctor-ibus*
      from   the-DAT.PL.   doctor-ABL.PL.
(10b) (Stolt 1964:140)
      bey   den   *universitat-ibus*
      at   the-DAT.PL.   university-ABL.PL.
(10c) (Stolt 1964:139)
      auf   alle   *loc-os*
      in   all-AKK.PL.   place-AKK.PL.

The Latin form can be interpreted either as an ablative or as a dative in (a)/(b) and it must be interpreted as an accusative in (c). The German preposition governs a dative noun in (a) and (b), and an accusative noun in (c), and therefore the case in a monolingual German sentence and that in a monolingual Latin sentence (*de doctoribus, in universitatibus, in locos*) are compatible. However, in the following examples, the German preposition requires the dative, but the noun receives the ablative, a case which is non-existent in German:

(11a) (Stolt 1964:140)
     in   der   *qualitate*
     in   the-DAT quality-ABL
     (Latin dative: *qualitati*, Latin equivalent: *in qualitate*)

(11b) bey   dem   *cult-u*
     at   the- DAT worship- ABL
     (Latin dative: *cultui*, Latin equivalent: *apud cultum, in cultu*)

(11c) vom   *praecaeptor-e*
     of-the- DAT   teacher- ABL *praecaeptore*)
     (Latin dative: *praecaeptori*, Latin equivalent: *de*)

Here, the Latin insertion is not only immune against matrix language influence (it does not receive German case marking), it even influences the matrix sentence itself. The German preposition is treated as if it was the equivalent Latin one, i.e. it now governs the ablative (as the corresponding Latin preposition would) instead of the dative which is required in monolingual German.[5] There is evidence then that the matrix language of this sentence is not monolingual (Early New High) German but rather a variety of German which is influenced by the Latin embedded language island.

There are other cases of German as the matrix language influenced by a Latin insertion. For instance, in the following example

(12) (608, Stolt 1964:71)
     obschon   *peccator-es*   izt   sind
     although   sinner-PL   now   be-3 PL
     'although they are now sinners'

the matrix language German requires a subject pronoun, while Latin doesn't:

German: *obschon sie Sünder sind*
Latin: *quamquam __ peccatores sunt*

The mixed sentence follows the Latin pattern, i.e. its syntax is not well-formed according to the matrix language German. Similar structures can be found and have been reported for other language pairs. For instance, in the Russian/Kazakh code-mixing corpus which will be the empirical focus of section 4, examples such as the following occur:

(13) (Muhamedova 2004, Kazakh/Russian, Almaty) [6]
      Moskva-da   КВАРТИР-У   име-ет
      Moskow-LOK   apartment-ACC   have-3P.Sg.PRES
      'he has an appartment in Moskow'

On the surface, this looks like a Russian sentence into which the Kazakh adverbial phrase *Moskva-da* is inserted. However, the matrix language is not grammatically well-formed in monolingual Russian, since Russian does not use the 'have'-construction to express possession but exclusively the locating copula construction:

(14) (fabricated example)
      У   иего   (есть) в   Москве   квартира
      With   him   (is)   in   Moscow   appartment

The Kazakh beginning of the code-mixed utterance (13), although 'just' an insertion, seems to impose on the sentence as a whole the syntax of Kazakh, despite the fact that all subsequent morphemes come from Russian. Equally, in the following example of Spanish-English code-mixing in New York reported by Angermeyer (2004:319),

(15) *Mira,   esos   son*   testers, *que*   had   to   use
       Look   these   are   testers which   I had   to   use

the matrix language of the subordinated sentence starting with *que* could be said to be English. However, its syntax is not well-formed in English where the subject pronoun cannot be dropped as in Spanish. If we want to look at the subordination as an insertion (see section 5 below), we are forced to conclude that the syntax of the matrix language is influenced by that of the embedded language.

In sum, we have seen that the inserted word or island can play an 'active' role in the clause, since it has an impact on the matrix language (the external syntax of the embedding).

4. Russian embedded language islands in Kazakh sentences

We now turn to a pair of languages (one a Turkic language, the other a Slavic/Indo-European language) of particular interest for testing morphological predictions on language mixing since the languages in contact both have rich morphological systems which nonetheless have a completely different make-up. While we have shown in the preceding section that the embedded language may have an 'undue' impact on the matrix language (and the matrix language an insufficient one on the embedding), we will now show that the matrix language can have an 'undue' impact on the embedded language. We will do so by looking at the internal grammar of embedded language islands.

Two differences between Russian and Kazakh nominal morphology are important in this context. One is very simple: Russian has a gender system, Kazakh does not. The second difference concerns the number system of the two languages. Both mark number in the nominal and verbal morphology (Singular/Plural), but Kazakh does so only on the head of a construction. The difference is most clearly visible in noun phrases containing an adjectival modifier: the Kazakh plural suffix /ler ~ lar/ is attached to the head, but not to the adjective, e.g. in *qazaq bala-lar-i* 'kazakh children,' while the Russian plural is marked on both the adjective and the noun, although often not by the same suffix, Russian being a highly syncretistic language, cf.: *высот-н-[ые]*PL.NOM. *дом-[a]*PL.NOM. 'skyscrapers'.

Against this background, consider the following examples that are typical for Russian embedded language islands consisting of a noun and an adjectival modifier in Kazakh matrices as produced by Kazakh-Russian bilinguals:

(16) уже   anau *стар-ый площадь*-tï   ne-ler-di   žönde-di
      already this old-???   square-???-ACC thing-PL-ACC   renovate-3SG
      'there the old square and so were already renovated'

(17) *высш-ий*   *школ*-dï bïtïr-d   osïnda (.) *транспорт-н-ый*
      highest-??? school-ACC finish-3PAST   here   traffic-SUFF-NOM.SG.MASK
      'after he had finished high school he worked here with the
      milicija[7]-da iste-dï.
      police-LOK work-3PAST
      traffic police'

(18) sosïn   ne (—) o orïs   *заведуюш-ий*-lar   ke-p
      then   uhm (—) R- Russian director-NOM.???.MASC-PL   come-CONV
      'then uhm when the Russian (library) directors came'

žňağï   *книж-н-ый*   *выставк-а*-lar   tema-lar-ï-n
this/uhm   book-BUFFER-???   exhibition-???-PL   theme-Pl-POS3P-AKK
ne-ler-i-n   sura-ğan-da,
thing-Pl-POS3P-ACC   ask-PART-LOC,
'and uhm asked for the topics for the book exhibitions and so on....'

In the first example, the embedded island *старый площадь* is integrated into the Kazakh sentence frame by a Kazakh late system morpheme *tï/dï* marking case. (We are clearly dealing with an island and not with two single word insertions as there is a relationship of dependency between the two words.) The noun *площадь* 'square' is inherently feminine in monolingual Russian; consequently, a modifying adjective needs to follow the feminine inflection as well: we would expect *старая площадь*. However, the speaker uses another suffix (*-ый*) which in monolingual Russian marks the nominative/accusative masculine singular. What has happened? One possible explanation would be that the noun has simply changed its gender from feminine to masculine:
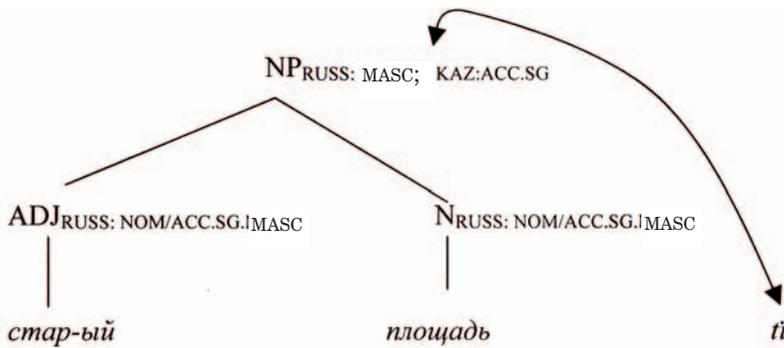
The NP as a whole receives its case marking from the Kazakh verb (see arrow) and, in fact, can be considered Kazakh. However, as outlined above, gender has no place in the grammar of Kazakh, and if the maximal projection of the N is Kazakh it remains mysterious what happens to gender marking on its way from the N up to the NP. Also, there is no reason why the noun should *change* gender in this particular language contact. Much more plausible is another interpretation: the noun has lost its gender marking altogether as it is integrated into a Kazakh matrix:
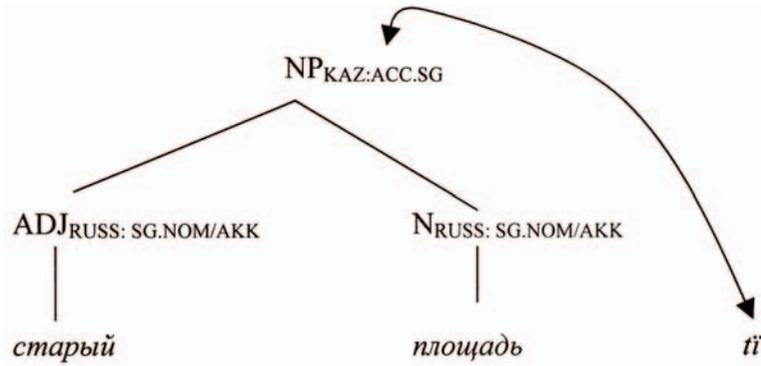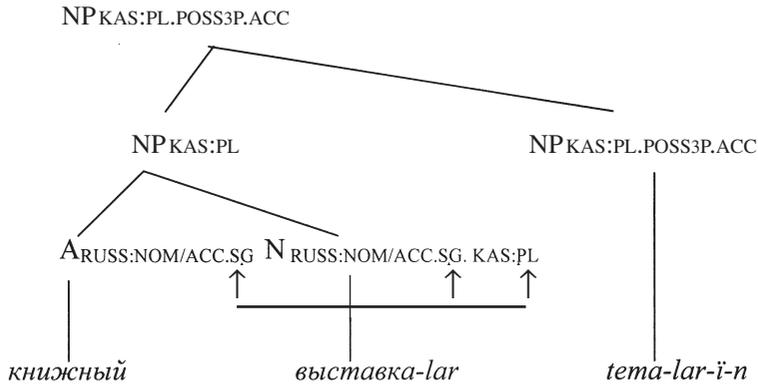
Under this interpretation, the suffix *-ый* has been reanalysed as a general marker for (genderless) adjectives. (We will discuss number marking below.) The Russian adjective is thus treated like a Kazakh adjective. This is compatible with the idea that the NP as a whole is treated like a Kazakh NP as well; however, it clearly contradicts Myers-Scotton's prediction that embedded language islands are under the exclusive regiment of the embedded language in their internal make-up.

The second example is very similar, but shows an additional aspect of how Kazakh can have an influence on the internal structure of Russian islands. *высш-ий школ-дï* once more is integrated into the Kazakh frame by the accusative suffix. Again, the noun is inherently feminine in monolingual Russian, but the adjective ends in the form *-ий* which is part of the masculine paradigm in monolingual Russian. The apparent conflict is resolved when we consider *-ий* to be a gender-neutral adjectival suffix. But in addition, the head noun in this example has lost its final vowel indicating number, gender and case (truncation): *школа* has become *школ*. (In monolingual Russian, the full NP would be *высш-а-ия школа*.) The truncation of the final /a/ indicating the feminine inflection is another regular phenomenon in the data under consideration. It is remarkable since the word-final syllable carries the main morphological information of the Russian noun (late system morphemes). Truncation provides further evidence for the hypothesis formulated above that inserted Russian is different from monolingual Russian in having lost the gender distinction.

The third example is slightly more complex. The embedded language island *книж-н-ый выставк-а* shows the same 'lack' of gender congruence as in the previous examples: an adjective which looks like a Russian masculine modifies a noun which looks like a Russian feminine.

Again, we suggest that the gender distinction has been lost, that the suffix *-ый* functions as a neutral adjective-forming suffix and that the *a*-suffix in the noun is no longer a carrier of morphological gender. But this time, the head noun in the embedded island (*выставка*) which is morphologically marked as singular in monolingual Russian is marked for number by the Kazakh plural suffix *-lar*, yielding a seemingly contradictory *выставка-lar* (Russian singular followed by Kazakh plural). In addition, the Russian adjective preceding the head is singular as well according to the grammar of monolingual Russian:

NP KAS:PL.POSS3P.ACC

NP KAS:PL   NP KAS:PL.POSS3P.ACC

A RUSS:NOM/ACC.SG   N RUSS:NOM/ACC.SG. KAS:PL

книжный   *выставка-lar*   *tema-lar-ï-n*

A straightforward interpretation is that bilingual Russian has lost the number distinction of monolingual Russian as well as the gender distinction. The syncretistic suffixes of Russian are no longer interpreted as carriers of number information. As a consequence, the noun may receive the Kazakh plural suffix, and the adjective remains unmarked for number as it is expected according to the Kazakh number marking system on the head noun. There is no real conflict between the Kazakh plural marking on the head noun and the form of the adjective: the type of Russian which is used for insertions does not mark gender or number, but has a generalised adjective suffix only. This in turn means that the morphological structure of the matrix language has been imposed on the structure of Russian.

Our interpretation is supported by single adjective insertions such as in the following example:

(19)   (Kazakh/Russian, Almaty)

частн-ый	*bır*	*närse-ler-ge*	*bar-ïp*	*ıste-y-d.*
private-???	one	thing-PL-DAT	go-CONV	work-PRES-3P

'They work in some private thing'

In this case, the Kazakh dative noun phrase contains a Russian adjective which is placed before the indefinite article, as it should be according to Kazakh word order. The head noun 'things' is marked as plural, but the adjective occurs in what in monolingual Russian sounds like a singular. Single adjectives are always inserted using this *-ый*-form in our data; we can now identify this ending as the neutral marker for adjectives, carrying neither gender nor number information. In fact, all adjectival insertions follow this pattern and exclusively occur in this form.

As shown in more detail in Muhamedova (2004), the large majority (76%) of Russian 'non-masculine' embedded language islands in Kazakh matrices are ill-formed according to monolingual Russian syntax. In the monolingual mode (i.e. when speaking Russian) the same speakers simplify the Russian gender/number system only 7% of the time. This shows that the starting point of the simplification must be bilingual, not monolingual Russian, although there is a certain spillover into the latter as well. Certainly it is not the case that the monolingual varieties have started to converge (cf. Myers-Scotton 1993:17), with the divergent patters in the code-mixed utterances being a mere reflex of this structural convergence.

We can only note in passing that similar simplifications are found in Uighur/Russian code-mixing (Muhamedova 2004:66-87) and are also reported in the literature for other Turkic/Russian contact situations such as Uzbek/Russian (Chamidova 1985), Kyrgyz/Russian (Krippes 1994) and Gagauz/Russian (Menz 1999). We are dealing with a widespread phenomenon of language contact here, not with individual, possibly idiosyncratic cases. In all of them, the structure of the matrix language has a strong impact on the structure of the embedded language islands.

5. Dependent clause constructions

The last observation we wish to make on the relationship between matrix and embedded language concerns clause-level subordination. We focus on the use of subordinators of one language which introduce subordinated structures in the other. This type of insertion is particularly interesting in language pairs in which clause-level subordination requires a grammatical pattern which is different from the one used in main clauses. For instance, German has V-late syntax in subordinated clauses but V2-syntax in main clauses, while Latin word order does not

differentiate between main and subordinated clauses. What happens when a German subordinator (complementiser) introduces a Latin clause? Does it lead to V-final syntax? On the other hand, does the insertion of a Latin subordinator (complementiser) into a German frame relinquish the V-late constraint in that language? [8] In the Early New High German/Latin data discussed in section 3, the conflict between German and Latin word order is so strong that code-mixing of this type is avoided altogether. Thus, while the language can change at the transition between main and subordinated clause, as in (20) (Latin subordinated clause, German main clause with V2) or (21) (Latin main clause with subordinated German object phrase, verb-late), it cannot occur after a subjunction, i.e. there are no examples such as (22):

(20) (Stolt 1964:195)
si	*enim*	*hoc*	*verum*	*esset*,		so	schiss		ich	dem
if	PART	that	true	be-COND.3SG.		then	shit-COND.1SG		I	the-DAT.SG.
'for if this was true, I would shit

pabst	auf	die	kron (218)
pope	on	the	crown
on the pope's crown'

(21) (Stolt 1964:186)
non	*cogitat*,		dass	Gott		etwas		hoher ist
non	think-3SG.PRES		that	God		something		higher is
'he does not think that God is something higher than a man'

denn	ein	mensch (484)
than	a	human

(22) (fabricated example)
* *non cogitat, quod* Gott ettwas hoher ist denn ein mensch

Here, the Latin complementiser does not match the German complementiser which would require V-late syntax. Latin conjunctions can only occur in German clauses when the corresponding German conjunction does not require verb-late syntax (i.e., in parataxis):

(23) (Stolt 1964:151)
et	*tamen*	haben		wir	ein	forteyl	fur		der	welt,
and	still	have-1PL		we	an	advantage	before		the	world,
'and still we have an advantage before the world,

quia	das	ampt	is	unser (510)
for	the	office	is	ours

since we have the office'

The German equivalent would be *denn* here, followed by a main clause as in Latin.

Russian and Kazakh differ even more radically in their treatment of subordination. While Russian has subordinated clauses introduced by subordinators and containing a finite verb, Kazakh dominantly incorporates subordinated clauses into main clauses through converbs, participles and deverbal nominalizers. Subordinating conjunctions are rare and restricted to certain types of subordinated clauses (such as *-sa* ('if-') clauses). Given these structural incompatibilities, it might be expected that Russian subordinators cannot introduce Kazakh subordinated clauses (marked by converbs, participles or deverbal nominalisers). However, this is not entirely true, since we regularly find examples such as the following:

(24) (Kazakh/Russian, Almaty)
частенько   *получает-ся,*   *что* (--)   aralastïr-a-mïz.
often   occur-REFL.3SG   that   mix-PRES-1PL
'Quite often it occurs that we mix the two languages.'

In (24), a Russian main clause is followed by the Russian subjunction *что* which introduces a subordinated complement phrase. This phrase is made up of Kazakh morphemes and would therefore seem to qualify as a Kazakh matrix into which the Russian subjunction is embedded. However, the construction contains a finite verb *aralastïr-a-mïz* and therefore violates the Kazakh pattern of nominal subordination. It seems that the use of Russian subordination has triggered a transition into main clause syntax according to the Russian monolingual model:

(25) (fabricated example)
köbinese   biz   (eki til-di)   aralastïr-a-mïz.
mostly   we   (two language-ACC)   mix-PRES-1PL

(25a) (fabricated example)
частенько получает-ся,   *что*   *мы*   *смеш-ива-ем (дба языка)*
often   occur-REFL.3SG   that   we   mix-IMPERF-PRES.1PL (two
                                                 languages)

This suggests that *aralastïr-a-mïz* should be treated as an embedded language island since its external syntax is governed by the preceding subordination.

The following example is similar:

(26)   (Kazakh/Russian, Almaty)

старейшины	*ѓород-а*	*все*	*так*	*ѓовор-яш*	sonda,	*что*	ol
oldest	town-GEN	all	so	say-3PL	then	that	he
žaňaǧï	so	žer-ge		bar-ïp	namaz	žasa-ğan	e-ken.
uhm this		place-DAT	go-CONV	prayer	do-CONV	AUX-EVID.PERF.3P	

‘All the elders of the town then said that he uhm had gone there and prayed there’

Here, the Russian subordinator *что* introduces a subordinated construction which is in itself complex. It contains a converb *-ïp* which is attached to the first verb ‘to go’ and a subsequent finite verb *eken* to which the converbial construction as a whole is subordinated. (In Russian, the complex subordinated construction would be rendered by a coordination.) *что* once more blocks the Kazakh subordination technique from being applied to *eken* as in the monolingual Kazakh equivalent:

(26a)	ol	žaňaǧï	so	žer-ge		bar-ïp	namaz	žasa-ğan
	he	uhm	this	place-DAT		go-CONV	prayer	do-CONV
	e-ken		dep	ayt-a-dï				
	AUX-EVID.PERF.3		CONV	say-PRES-3				

Finally, consider the following case of a complementiser phrase introduced by Russian *что*:

(27)   (Russian/Kazakh, Almaty)

endi (.)	*единственный*	*что*	*они*	*вот*	äli	üyren-be-gen	ǧoy
well (.)	the-only-ADJ	that	they	so	yet	learn-NEG-PERF	PART

‘The only thing was that they had not yet learned it’

The main clause is Russian,[9] and so is the subjunction which introduces the dependent clause. The dependent clause itself contains a Russian subject pronoun in the plural, but the finite perfective verb is Kazakh.

In all three examples, the Russian subordinator has an impact on the Kazakh subordinated construction. This impact consists of blocking the nominal Kazakh subordination technique which is replaced by a finite construction like in Russian. However, the construction is well-formed according to Kazakh main clause syntax. The matrix language is hard to determine. The fact that the subordinator determines the choice between finite and non-finite constructions is evidence for Russian being the matrix language. On the other hand, the Kazakh finite construction

could be interpreted as a simple main clause which is preceded by a Russian subjunction which has been reanalysed as a subordinating particle.

While the subordinated clausal construction was a complementiser clause in object or subject position in the last three examples, there are also Kazakh clauses introduced by the Russian subordination *чтод/чтоды* expressing purpose. Here is an example:

(28) (Kazakh/Russian, Almaty)

o-lar	*ТОЖЕ*	*заинтересованы*	*чтоды*	adam	otïr-sïn.
they-PL	also	interested	that(FINAL)	man	sit-3PL.OPTATIV

'they are also interested that people sit down (in the bus, i.e., use public transport)'

In (28), the Kazakh clause is clearly finite, and therefore once more appears in the format of a Kazakh main clause despite the fact that it is subordinated to a Russian main clause by the subjunction *чтод*. However, other than in ex. (25-27), the Kazakh clause is not well-formed in this case. The finite verb *otïr* receives the optative suffix, despite the fact that neither Kazakh nor Russian allow an optative in the respective monolingual version of the sentence:

(28a) (monolingual Russian, fabricated example)

Они	*тоже*	*заинтересованы*	*в*	*том,*	*чтоды*	*люди*
they	also	interested	in	that	that(FINAL)	people

езди-л-и	*(в этих автодусах)*
go-PRET-PL	(in these buses)

(28b) (monolingual Kazakh, fabricated example)

olar	el-diň	aftobus-qa	otïr-ǧan-ï-n	qala-y-dï
they	people-GEN	bus-DAT	sit-PART-POSS.3PL-ACC	want-PRES-3PL

Since monolingual Russian requires a past form after *чтобы*, and Kazakh a participle in the respective nominal construction, the grammar of the code-mixing does not conform to either 'code.' Clearly, the speaker has chosen a new format which combines the Russian subjunction with a Kazakh form which fits its meaning (optative in purpose constructions).

In this section, we have looked at clause-level subordination in mixed utterances. Although the two languages in contact diverge quite radically in the way they express clause-level subordination, mixed utterances with a Russian subjunction (complementiser) fol-

lowed by a Kazakh clause are frequent in our data. The resulting construction certainly cannot be analysed as a case of a simple lexical insertion of a Russian subjunction into a Kazakh matrix frame (as suggested in Myers-Scotton 2002:2u-12). Since the subjunction triggers main clause syntax, it imposes a 'Russian' type of clausal subordination on Kazakh. It is difficult if not impossible (in (25)) to identify the matrix and the embedded language, especially if the clause (complementiser phrase) is taken as the unit of analysis.

6. Conclusion

In this paper, we have argued that a neat separation between matrix and embedded language is impossible. Arguing on the basis of Kazakh/Russian as well as Latin/Early New High German materials we have shown that embedded language islands may be under the regiment of the matrix language, and that the matrix language may be influenced by the embedded language. We also showed that mixed subordinated clauses occur which make it impossible to identify the matrix language.

We have taken Myers-Scotton's model as a point of departure for our analysis, since it is the most explicit, most elaborate and most cited model in this field. This model has been criticised elsewhere (cf. lately Berruto 2005). Our intention was more general, though. We wanted to argue for an approach to code-mixing which takes the syntactic structure of the mixed utterance as the starting point, rather than the monolingual 'codes' which these mixed utterances seem to refer to. Our examples demonstrate that often, there is no monolingual code which can be taken as the point of reference. This conclusion is also reached by Myers-Scotton in her 2002 theory with respect to the matrix language; here she insists that the matrix is not identical to any single 'monolingual' language but is just an abstract construct. However, this has far-reaching consequences. For instance, it is unclear how notions such as the 'congruence' between the matrix and the embedded language (the lack of which is made responsible for the occurrence of islands in her theory) should be defined. It is even unclear how morpheme order and late system morphology can be established which in turn are essential for the definition of the matrix language: if the latter cannot be equated with the monolingual code, we have no means of defining it. The conclusion, however, is inevitable: bilingual talk cannot be analysed as a mixture of two monolingual codes. An alternative has been formulated in 1998 by

Sebba in a paper 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: "Congruent categories are categories of the grammar which are treated as 'the same' in L1 and L2 by bilinguals" (Sebba 1998). We believe that this is the right way to proceed.

Address of the Authors:

Peter Auer (Corresponding Author), Deutsches Seminar I, Universität Freiburg, D-79085 Freiburg, Germany <peter.auer@germanistik.uni-freiburg.de>

Raihan Muhamedova, The German-Kazakh University Almaty Kazakhstan, ul. Pushkina 111/113, 050010 Almaty, Kazakhstan
<rmuhamedowa@yahoo.de>

Notes

[1]   In this paper, we will restrict ourselves to 'clause-internal' mixing, with the clause being identical to what is often called a Complementiser Phrase in Generative Grammar. The argument can also be made for 'sentences.' Of course the segmentation of spontaneous (conversational) bilingual speech into 'sentences' or 'clauses' is not a trivial issue (cf. Auer 1999b).

[2]   For a discussion of the matrix language and how it can be defined, cf. Boumans 1998, Ch. 2. Myers-Scotton's view contrasts with other, more verb-centred approaches.

[3]   Cf. Stolt 1964:15-47 on details of the available records and their representativity for the bilingualism of 16th century intellectuals.

[4]   There is of course no space to summarise her theory here. The following discussion reflects our own analysis of the data.

[5]   As Stolt (1964:141) points out, the German prepositions are not equally susceptible to this embedded language influence. While the German dative seems to be replaced quite regularly by the ablative, the accusative is more resistent. For instance, in the prepositional phrase *durch einen latron-em* (through a-ACC.MASC.SG. thief-ACC.MASC.SG.), the accusative case required by German durch is realised by the Latin accusative. The equivalent Latin phrase would have no preposition but only an ablative suffix (*latron-e*).

[6]   In this and the subsequent examples for Kazakh/Russian code-mixing, Russian is transcribed in cyrillic alphabet while Kazakh is transliterated into Latin alphabet.

[7]   *Milicija* is an established loan word in Kazakh.

[8]   A discussion of this issue with regard to Germanic languages in contact with English or French can be found in Treffers-Daller (1994:189-201).

[9]   The adjective *единственным* is ungrammatical in monolingual Russian (where a neuter form is required: *единственное*) but follows the pattern of bilingual Russian described in the previous section, i.e. it is neutralised with regard to gender and takes the unmarked masculine suffix.

Bibliography

ANGERMAYER Philipp (2005), *'Speak English or what?' - Codeswitching and interpreter use in New York Small Claims Courts*, Ph.D. Thesis, New York University, Department of Linguistics.

AUER Peter (1999a), "From codeswitching via language mixing to fused lects: Toward a dynamic typology of biligual speech", *International Journal of Bilingualism* 3: 309-332.

AUER Peter (1999b), "How should we and how can we determine the 'base language' of a bilingual conversation?", *Estudios de sociolingüistica* 1: 129-144.

BACKUS Ad (1996), *Two in one: Bilingual speech of Turkish immigrants in the Netherlands*, Tilburg, University Press.

BERRUTO Gaetano (2005), "Dialect-standard convergence, mixing, and models of language contact : The case of Italy", in: Peter AUER, Frans HINSKENS & Paul KERSWILL (eds.), *Dialect Change*, Cambridge, Cambridge University Press: 81-95.

BOUMANS Louis (1998), *The Syntax of Codeswitching. Analysing Moroccan Arabic/Dutch conversation*, Ph.D. Thesis, University of Nijmegen.

CHAMIDOVA S.M. (1985) - Хамидова, С.М., Русские словарные единицы в узбекской разговорной речи (к постановке вопроса). *Советская тюркология* 1: 20-29.

KRIPPES Karl A. (1994), "Russian code-switching in colloquial Uzbek, Kazakh and Kyrgyz", *General Linguistics* 34: 139-144.

MENZ Astrid (1999), *Gagausische Syntax. Eine Studie zum kontaktin-duzierten Sprachwandel*, (Turcologica 41), Wiesbaden, Harrassowitz.

MUHAMEDOVA Raihan (2004), *Untersuchungen zum Kasachisch-Russischen Code-Mixing (mit Ausblicken auf den uigurisch-russischen Sprachkontakt)*, Unpublished Ph.D. Thesis, University of Freiburg, Germany.

MUYSKEN Pieter (2000), *Bilingual Speech: A Typology of Code-Mixing*, Cambridge, Cambridge University Press.

MYERS-SCOTTON Carol (1993), *Duelling Languages*, Oxford, Clarendon.

MYERS-SCOTTON Carol (2002), *Contact Linguistics*, Oxford, Oxford University Press.

PAUL Hermann (18983), *Principien der Sprachgeschichte*, Halle, Niemeyer.

SEBBA Mark (1998), "A congruence approach to the syntax of code-switching", *International Journal of Bilingualism* 2: 1-10.

STOLT Birgit (1964), *Die Sprachmischung in Luthers Tischreden*, Stockholm, Almqvist & Wiksell.

TREFFERS-DALLER Janine (1994), *Mixing two languages. French-Dutch contact in a comparative perspective*, Berlin, Mouton de Gruyter.

TÜRKER Emel (2000), *Turkish-Norwegian Codeswitching*, Universitetet i Oslo, Faculty of Arts.