Towards a Turn-based Categorization of Codeswitching and Codemixing

Journal Article

Heike Havermeier

This article addresses a terminological issue of categorizing bilingual speech. Previous categorizations often fail to contribute suitable categories for describing the form of codeswitching in oral conversations, since none of them considers the characteristic features of spoken language use. Therefore, a new categorization and terminology is suggested. This employs turn construction units, the main organizational parameter in spoken language conversations, as a criterion to distinguish two main types of codeswitching. Codeswitching inside turn construction units not affecting the base language is called CODEMIXING. A change of the base language at the outlines of these units is called CODEBREAKING. Codemixing is further distinguished into *free* and *bound codemixing*, depending on whether elements from two languages are syntactically dependent on each other, or are syntactically free or self-contained. It can be shown that the types of bilingual speech classified according to these criteria are deployed considerably differently in a German-Swedish corpus.

1.   Introduction

When describing bilingual language use, research has always reflected the need to classify the incidences of codeswitching according to their form, both for the purpose of theory formation and for the practical purpose of describing and analyzing bilingual corpora. Since the first approaches in the 1980s (e.g., Poplack 1980; Poplack & Sankoff 1988), a number of description and categorization models have been introduced. Among those, the MLF-Model presented by Myers-Scotton (1993, 2002) and the codemixing-typology presented by Muysken (2000, 2013) have had the greatest impact on current research. However, none of these models is without controversy. Both Poplack's and Myers-Scotton's models have been rejected by a majority of researchers, at least in their universality claims, since findings in various corpora provide counterexamples to the basic assumptions of those models, as well as to the suitability of their categorization categories (see, e.g., Treffers-Daller 1997: 178; Clyne 2003: 191f; Edel 2007: 47). Muysken's typology is currently the most accepted categorization, but is not unquestioned either (see, e.g., Bhatt 2013). The main point of criticism is that, in his more recent works, Muysken suggests that only one type of bilingual speech could be customary in a specific language contact situation (cf. Muysken 2013: 714).

This paper aims to highlight another aspect that was found to be an issue when looking for categories suitable to describe a spoken language corpus containing bilingual German-Swedish conversations: None of these models take into account that bilingual speech is actually speech, i.e. oral language use with all its patterns and practices deviating from the norms and properties of written language. Instead, research on codeswitching so far investigated spoken language as if it were the same as written language.[1] The uncritical usage of written language rules and concepts for bilingual speech has been criticized before, e.g. by Gardner-Chloros, who argues that

---

[1]   Interestingly enough, codeswitching in written language has likewise been ignored by the same research tradition, with the result that written multilingual discourse is an "under-researched" (Sebba 2012:1) topic and has been established as an own field of study only recently (e.g., Sebba et al. 2012).

we need to look outside language as a closed system, which has been the focus of research in the structuralist tradition since Saussure and Chomsky, just as pragmaticians and discourse analysts have done so successfully in the last few decades (Gardner-Chloros 2009: 106).

Still, a suggestion for more conversation-oriented analysis tools remains a desideratum.

This paper aims to introduce a new model to categorize and describe codeswitching based on relevant categories in spoken language. In the first step, it includes turn construction units, and in the second step syntactical government combined with agreement, which are especially relevant in the Germanic languages since the model has been developed based on these.

In the following, Muysken's typology is described in more detail, since the suggested new categorization and terminology is mostly built on his work. The essential terminology of spoken language research going back to conversational analysis is explained as well. After a brief presentation of the corpus data which the observations are based on, I present the suggested model, which is illustrated with corpus examples. The corpus analysis shows that incidences of bilingual speech categorized according to the suggested criteria serve different functions and occur in different conversation scenarios. Moreover, there are differences in the further implementation of inserted elements.

2. Previous Segmentation Categories for Bilingual Speech and Speech in General

Muysken's typology is based on the main distinction between intersentential and intrasentential phenomena. In his initial model presented in 2000, these categories are called *codeswitching* and *codemixing*, whereby codemixing is defined as "cases where lexical items and grammatical features from two languages appear in one sentence" (Muysken 2000: 1). The term *codeswitching* is, however, widely used in the research literature as an umbrella term for all instances of speech or texts in which two or more languages contribute with lexical matter (see, e.g., Myers-Scotton 2002: 22; Gafaranga 2007: 279; Backus & Dorleijn 2009: 76; Gardner-Chloros 2009: 13). Due to that, Muysken uses the term synonymously with *codemixing* or as a super-ordinate term in his more recent works (cf. Muysken 2013: 710). Also in this paper, CODESWITCHING will be used for all instances in which matter from more than one language occurs in one conversation (terms defined in this paper will be marked with SMALL CAPITALS in the following).

Muysken distinguishes different processes concerning intrasentential CODE-SWITCHING: The first one is *insertion*, which means that items from one language are embedded into constructions in the other language. These items can be of different length, from single lexical items to larger constituents, as long as their use does not affect the matrix language of the whole sentence (cf. Muysken 2000: 3f and 61ff). If the matrix language is changed at a linear point inside a sentence, this is categorized as a different process, called *alternation* (Muysken 2000: 96). However, even syntactically relatively loose elements such as adverbs and discourse particles are considered as alternation, at least in the initial typology presented in 2000 (cf. Muysken 2000: 97f). In a revised version, the switching of discourse particles is assigned a separate category, *backflagging*, at least when they are employed to signal the original community language (cf. Muysken 2013: 713f). The last

type of codemixing according to Muysken's categorization is *congruent lexicalization*. These are instances of bilingual speech in which both involved languages contribute lexical matter and grammatical structures, typically without a clear linear switching point. This often goes hand in hand with structural convergence and generally affects languages or varieties that already are very similar, so it is sometimes not possible to tell which one accounts for which element or structure in the utterance (Muysken 2000: 122f).

Some aspects of this classification are controversial (and have been criticized before, see, e.g., Edel 2007: 91; Bhatt 2013). One point of critique is that some types, such as insertion, are defined by their form, whereas others, especially backflagging, are defined by their motivations and causes. Another point is that it is not convincing that items of particular word classes, such as adverbs, are suggested not to be insertions, although they are clearly nested in speech formulated in another language. However, Muysken's most basic distinction between intrasentential code-mixing and inter-sentential codeswitching has seldom been questioned, even though this definition is problematic. First of all, Muysken does not give any definition of *sentence* or *intra-sentential*. However, a *sentence* is not a well-defined size, at least not in oral language use,[2] the typical context of bilingual speech and the context that hitherto has mostly been accounted by codeswitching research (cf. Sebba 2012). In spoken language, there are no identifiable punctuation marks to flag boundaries of an utterance, and they could not simply be inserted as in a written text. As Auer (2009) remarks, speech is marked by time pressure along with simultaneous production and reception, which determines an absolute linearity. These demands result in syntactic structures that differ from those of written language (cf. Auer 2009: 8ff). Spoken language contains many construc-tions which do not represent full clauses. Some of those instances are the result of interruptions (both self-interruptions for the purpose of self-repairs as well as interruptions by interlocutors and other external forces). They might, therefore, be rejected as exceptional mistakes. However, the majority of utterances not representing full clauses contain exactly what the speaker wanted to say. Many utterances consist of only a single word or phrase, e.g. as an answer to a question. Furthermore, there are routine formulations that may be classified as grammatically incomplete from a written language point of view, but are nonetheless meaningful (cf. Günthner 2007: 76). Such formulations can also be found in the corpus, for instance in Example 1. The utterance starts with two formulations that show syntactic peculiarities.

The first one, "aber um jetzt noch mal" ('but to now once again'), is built up like an infinitive phrase, but does not contain an infinitive at the end. The formulations can be interpreted as an ellipsis, assuming that the envisaged formulation is "um jetzt noch mal auf x zurückzukommen" ('to get back to topic x') and that the last words are skipped.

The second one, "weil wenn du sagst schwedisch" ('because when you say Swedish') has the form of a subordinate clause, without having any superordinate main clause. However, the formulation works fine on its own, since it is quite common in German to utter subordinate clauses of the type "now that you say x" or "when you say x" without a main clause to raise a certain topic.

---

[2] The following discussion on the syntactical struc-tures in oral language use is mainly based on work on German, because my analysis concerns a pre-dominantly German corpus. Still, many of the considerations also hold for other languages.

Given that not even clauses are a set size in this context, it seems absurd to demand the identification of sentences. This does not mean that there are no full clauses and well delineated sentences in oral speech, but that they are not the topmost organizational parameter in spoken language conversations (cf. Duden 2005: 1225 and 1238; Günthner 2007: 76).

To find a segmentation category that fits better for bilingual speech, it is advisable to have a look at the categories that other lines of research on oral language use rely on. Linguistic work on conversations, as conversation analysis going back to Sacks et al. (1974), usually builds upon *turns* and *turn construction units (TCU)*. These are orientated on information units and play a crucial role in negotiating turn-taking between interlocutors. One turn may contain more than one TCU, but their outlines constitute possible breaking points called *transition relevance*

---

[3] In the examples in this paper, Swedish is represented in *italics* and German is represented in normal typeface, in the transcriptions as well as in the translations. An explanation of the signs for non- and para-verbal elements in the transcripts can be found at the end of this paper, along with an explanation of the abbreviations in the glossed lines according to the Leipzig Glossing Rules.

Example 1[3]

```
(1)A: aber um jetzt noch mal <laughs> weil wenn du sagst schwedisch↓
       but  to  now  again once          because when you.SG say   Swedish
       'But to now once more (laughs) because when you say Swedish.'

(2)     Ich hab ich weiß nicht ob euch das auch so geht
         I   have I  know not  if you.PL that also likewise go
       'I have – I do not know how you feel about it –'

(3)     ich hab aber immer problem mit Öron och Ögon↓
         I   habe though always problem with ear  and  eye
       'but I have always problems with ear and eye.'
```

Example 2

```
(1) A: äh die frage ist aber jetzt↓ für mich auch↓ also äh
       er the question is though now  for  me   also  so   er
       ,Er, but now the question is also for me, well, er,

(2)    ob man die jetzt (-) REIN fürs deutsche jetzt macht oder
       if INDF them now      mere for+the German    now  make  or
       'if you now plan them exclusively for German or

(3)    für TYska macht↑  für ÄMnet  tyska↑ (-) oder die   gleichzeitig
       for German makes  for subject-DEF.N.SG German or   them simultaneously
       'plan them for German, for the discipline German studies, or if you likewise

(4)    auch offen(.) hält(.) für= oder UND offenhält(.)
       also open     holds    for  or   and openholds
       'also keep them open for- and keep them open

(5)    für andere↓ teilnehmer↓
       for other/others participants
       'for other(s) participants.'
```

places (*TRP*). At TRPs, another interlocutor can take over without interrupting the speaker. TCUs can be of different length, from single words (for example discourse particles) to longer narrations. Their outlines are primarily constituted by prosody. Even though clausal and sentential structures also play a role in the constitution of TCUs, their outlines do not always match. In a revised description of TCUs and their properties, Selting (2000: 504) states: "In one set of cases, prosody seems to override syntax; in the other, syntax seems to override prosody". An example of a prosodic unit subsuming several clauses can be found in Example 1 above. Despite the fact that there are several syntactical breaking points before, the prosody is held high from the beginning of line (2) until it signals a closing after *ögon*. The opposite case can be seen in Example 2.

In this example, the speaker repeatedly signals by means of the prosody that he has finished his point, especially after the second *offenhält* ('keep open'), after *andere* ('other, others') and after *Teilnehmer* ('participants') in lines (4) and (5). What is interesting is that, after each of these TRPs, a syntactical construction would be completed; however, at the same time, this construction is upheld by the speaker. When none of the other interlocutors responds, the speaker continuously adds more complements – which do not contribute much to the meaning –, first to the clause and then to the prepositional phrase. Thereby, the word *andere* in line (5) even changes its function from pronoun to adjective during the formulation process. This gradualness and flexibility is an essential difference to written language, where readers are presented with a sentence as a complete, closed unit.

Given that research on spoken language conversation has accounted for sentences not being a suitable topmost segmentation category for speech, it also does not seem appropriate to use sentences as topmost categorization criteria for bilingual speech, either. I therefore suggest using TCUs as categorization criteria also for bilingual speech phenomena. In the following, it will be shown what this categorization contributes to the analyses of a bilingual corpus.

3.   Data

The observations in this paper are based on a corpus consisting of German-Swedish conversations, which was compiled in the context of a dissertation project on multilingual practices in academic working environments (Havermeier 2015). The participants are staff members of Swedish universities working in German studies. Both Swedish and German are frequently used in their work places. The participants are fluent speakers of both languages, though they are L2-speakers of one or the other. The corpus contains about 40 hours of conversations which were conducted between the participants as well as between participants and their students. The recordings took place in various situations of the participants' everyday working life. Due to that, the corpus is divided into subcorpora representing different social situations.

The corpus has been analyzed with respect to the usage of more than one language in a conversation. Passages containing CODESWITCHING have been transcribed and the transcription conventions are based on the GAT system (Selting et al. 1998, see also the explanations for prosodic signs at the end of this paper).

4.   Codemixing and Codebreaking

4.1   Distinguishing Codeswitching Based on Turn Construction Units and Base Language

When CODESWITCHING is distinguished into different categories based on whether the switch takes place at the outline of a TCU or inside a TCU, the analysis differs considerably from a categorization based on the concepts *intersentential* and *intrasentential*, even in instances where there are complete, clearly delineated clauses. When applied to examples from the German-Swedish corpus, the newly suggested categorization would subsume Examples 3 and 4 under one category: Although there is a codeswitch inside a clause in Example 1, whereas the CODESWITCHING takes place at the outlines of clauses in 4, German and Swedish are used inside a turn construction unit in both cases. In contrast, Example 5 belongs to another category, where the languages are used in different turns.

  A relevant factor for this categorization is also the conversational *base language*, a concept that has already been applied e.g. by Treffers-Daller (1997), based on Grosjean's

Example 3

```
(1)A: also wir haben auch so ein mall↓
       PART  we   have  also  such a template
       'Well, we also have such a template.'

(2)B: für power points↑
       for  power point.PL
       'For power point presentations?'

(3)E: ja ja
       yes yes
       'Oh yes'
```

Example 4

```
(1)A: er sagt ja det är kvalitetshöjning↓ det gör vi↓
       he  says PART that is  quality increase      that do  we
       'He says: That is quality improvement. Let's do that.

(2)B: hm↓ ja↓
       uhm  ja
       ,Um, ok.'

(3)A: also es gibt eigentlich ja dafür kein GELD↑(-)aber er sagt jA↓
       so   it  gives actually      PART for that  no  money    but  he  says yes
       'So, normally, there is no money for that, but he says yes.'
```

(1995) model of bilingual language modes, but has not yet been incorporated in general classification models. The basic idea is that, even in bilingual speech, there is most often one language serving as the overall base of communication. Elements from the other language can be inserted into it, or an overall change in the base language can take place

at a certain point. The distinction between two languages contributing with different amounts of influence is thus similar to that between *matrix language* and *embedded language* in the MLF-model (cf. Myers-Scotton 2002: 14), which is more widely spread in the research literature (see, e.g., Muysken 2000: 64). The concept of a *matrix language* is based on the assumption that one of the languages is always in charge for the syntactical formulation of bilingual speech. Muysken also mentions the concept of a discourse-oriented base language briefly, but does not elaborate on it, since it does not contribute to his sentence-oriented pers-pective (cf. Muysken 2000: 64f).

Applied to the corpus examples given above, we observe that, in Example 5, the base language changes between the turns in (6) and (7). In Example 3 and 4, the base language of the conversation is and remains German. Although Swedish appears as well, it is embedded and does not affect the language chosen for the following turns.

In terms of terminology, I suggest to adopt CODEMIXING for bilingual speech inside a TCU that does not change the base language. This is in conformity with Muysken and other authors who use *codemixing* as a label for embedded CODESWITCHING, often,

Example 5

(1) A: oder= warten wir noch n moment↑
         or   wait   we   still   a moment
         'Or – should we still wait a moment?'

(2) B: (Name)[1]   kommt doch noch oder↑
         (person's name) comes   but   also or
         'But N.N. is also coming, right?'

(3) A: doch oder↑
         yes   or
         'Yes, right?'

(4) C: ja↓
         yes
         'Yes.'

(5) A: wer kommt denn ah ja (Name) kommt ja noch↓
         who comes Q   oh yes (person's name) comes PART still
         'But who else is going to come? Oh, yes, N.N. is coming as well.'

(6) B: ja ich will noch=auch(.) so↓
         yes I want still also   so
         'Well, I also wanted to… ok.'

(7) A: *ja vi= vi har förberett oss på= på svenska↓* <laughs>
         yes we= we have prepared   REFL on= on Swedish
         'Yes we… we have prepared this in- in Swedish.'

(8) D: <in joking voice> *det gläder mej↓*
                              that gladdens me
         'I'm glad to hear that.'

but not always in contrast to *codeswitching* (see, e.g., Auer 1999; Dirim & Auer 2004;[4] Kallmeyer et al. 2002; Matras 2009: 101). Nevertheless, I adhere to the usage of CODE-SWITCHING not as an opposite, but as the general term for both (and other) types, since it is widely established as a comprehensive term and used more or less identical with *bilingual speech*, even as a label for handbooks etc. (see, e.g., Bullock & Toribio 2009; Gardner-Chloros 2009; and even Muysken 2013).

For the type of CODESWITCHING which entails a change of base language, there is no equally suitable established expression. The term *alternation* describes a similar process, but is clearly associated with a break inside a syntactical construction, not at the outline of a unit such as a TCU. The only terminology highlighting the same difference of the processes distinguished here is Clyne's *trans-*

---

4 According the terminology introduced by Auer (1999), the distinction between *codeswitching* and *language mixing* (in subsequent works: *codemixing*; cf. Dirim & Auer 2004: 158) is not primarily based on the form of codeswitching, but rather on its functions together with its social and communicative background (cf. Auer 1999: 310). This use of the terminological pair is rather common in research on German as a contact language, whereas a form-based understanding of *codemixing* is more common internationally.

Example 6

```
(1)   wenn ich sage er IST krank↑ (2) då har vi det glasklart↓ va↑ (1)
      when  I  say  he  is   sick      then have we that crystalclear    Q
      'If I say he is sick? (Pause) Then we have it crystal clear, right?
```

```
(2)   då Ä han det↓ då är det FAKtum↓ men↓ om jag säger er ist WOHL krank↑(2)
      then is he  that then is it  fact   but  if  I   say  he is probably sick
      'Then that's what he is. Then it is a fact. But: if I say he is probably sick?
```

```
(3)   <<performing surprised> han är antagligen↓ han måste vara sjuk↑ (1)
                              he is  probably   he must  be   sick
      (performing surprised) 'He is probably. He has to be sick.
```

```
(4)   er ist wohl krank↓ > (5) så det är han är↓ ju↓ möjligen↓
      he  is probably sick     so that is he is  PART  possibly
      'He is probably sick. (pause) So, that means he is well perhaps
```

```
(5)   <<fast> kanske också↓>  <<performing> han är ju möjligen sjuk
              maybe  also                   he is PART possibly  sick
      '(fast) maybe also. (performing surprised:) He is perhaps sick
```

```
(6)   därför är han inte här↓ (1) er ist wohl krank↓ > (3) <takes a deep breath>
      therefore is he  not here      he is probably sick
      'therefore he is not here. (pause) He is probably sick. (pause, takes a deep breath)
```

```
(7)   + deshalb haben wir also diese äh= modalwörter↓ aber die können lei= a=auch
      therefore  have  we  thus these er  modalwords  but  they can unnfort=   also
      'Thus, that's why we have those, er, modal words. But they can unfort- a- also
```

```
(8)   son bisschen verwechselt werden↓ das tun übrigens die deutschen auch
      such+a  bit    mixedup     become that do  besides the Germans    also
      'somewhat be mixed up. That happens to Germans also, by the way.
```

```
(9)   hier angeblich und anscheinend↓ det är immer immer svårt att hålla isär↓
      here allegedly  and probably    that is always always difficult to hold  apart
      'Those ones, allegedly and probably. That is always always hard to keep apart.'
```

version, defined as "a crossing over from one language to another rather than a transference of an item, feature or construction" (Clyne 2003: 80). However, this term did not gain further currency. I therefore suggest the new term, CODEBREAKING, for CODESWITCHING affecting the communicative base language beyond the current turn construction unit.

It is important to observe that this categorization is only applicable to code-switching when there is a clear base language. This does not hold for all instances of bilingual speech. There is, for example, no base language recognizable when both languages over several construction units contribute to the speech production in the way that Muysken describes as *congruent lexicalization*. The same applies when there are repeatedly switches over several TCUs. Passages of that kind can also be found in my corpus, for instance in Example 6. The rather monological passage is found in a discussion about the translation of a text, specifically the sentence "Er ist wohl krank" ('He is pro-bably sick.'). Due to constant naming of both the original and possible translations in both languages, the speaker begins to switch between languages for metalinguistic comments as well (for a further description

of this codeswitching pattern see Havermeier 2015: 223-232).

Thus, CODEMIXING and CODEBREAKING are not fitting for all instances of bilingual language use, but that does not query the suitability of the suggested classification model altogether. It just shows that we need different description tools for different instances of conversation. For my model, it means that a third category besides CODEMIXING and CODEBREAKING is required, namely CODESWITCHING without a base language. The codeswitching patterns falling into that category can be further classified as well, for example with Muysken's concepts *alternation* and *congruent lexicalization*. However, the focus of this paper lies on instances of bilingual speech where CODE-MIXING and CODEBREAKING are applicable. Thus, I will not go further into other forms, but instead demonstrate what implications the classification into CODEMIXING and CODEBREAKING has on the corpus analysis.

4.2 Deployment of Codemixing and Codebreaking in Bilingual Conversations

An analysis of CODESWITCHING according to CODEMIXING and CODEBREAKING shows that incidences belonging to these categories

have indeed more features in common than those determined by the definition of the terms. As described above, the corpus of German-Swedish conversations is divided into subcorpora containing different social situations typical for academic working life. It covers academic discussions (seminars, conference talks, etc.), administrative conversations, and socializing conversations (small talk, lunchroom conversations) between colleagues as well as communication with students when teaching classes and in consultations. As shown in Table 1, the appearance of CODEMIXING and CODEBREAKING is not evenly distributed among these conversational situations.

The table shows both the total number of occurrences (on the left-hand side of each column) and the percentage contribution of each type (on the right-hand side of each column). The total figures might be delusive due to the very different sizes of the subcorpora (see recorded hours). However, the percentage values give a clearer picture. CODEMIXING is by far the most preferred type of bilingual speech in these situations, or rather for this speaker community. It seems nearly the only common pattern in conversations with administrative or academic matters.

CODEBREAKING is, however, not uncommon in student consultations, in teaching, nor, to a certain degree, in socializing conversations. Teaching also contains a certain amount of CODESWITCHING without a base language. That is due to passages of collective translation work, as shown in Example 6 above. Additionally, CODESWITCHING can be triggered by the fact that students sometimes use another language due to convenience (see also Söderlundh 2012).

The differences in numbers of CODEMIXING or CODEBREAKING are thus not directly caused by the factors that define the situation – that is, topic and purpose of the conversation – but rather by factors such as number of and hierarchy between the interlocutors. These lead to different communicative needs, which for their part are the motive for CODESWITCHING. CODESWITCHING then takes the form that is either most suitable for this function, or is common for this function in the given speaker community. Table 2 gives an overview of the motives and triggers for CODESWITCHING in the corpus, in relation to the form-based categories.

Since the motivation for and functions of CODESWITCHING are beyond the scope of this article, the categories in the first column

	recorded hours	CODEBREAKING		CODEMIXING		no base language		total	
Administrative	6.5 h	2	1.1 %	174	98.9 %	0	0.0 %	176	100 %
Socializing	1.8 h	7	14.6 %	41	85.4 %	0	0.0 %	48	100 %
Academic	11.3 h	4	1.7 %	226	97.4 %	2	0.9 %	232	100 %
Student consultion	0.3 h	4	50.0 %	4	50.0 %	0	0.0 %	8	100 %
Teaching	17.8 h	49	8.1 %	418	69.2 %	137	22.7 %	604	100 %
total	37.7 h	66	6.2 %	863	80.8 %	139	13.0 %	1068	100 %

Table 1 Distribution of CODEMIXING and CODEBREAKING in Different Social Situations

	CODEBREAKING		CODEMIXING		no base language		total	
Subject to metalinguistic comment	4	0.8 %	463	94.9 %	21	4.3 %	488	100 %
Reported speech/quotes	4	2.0 %	172	87.8 %	20	10.2 %	196	100 %
Language domains	1	0.6 %	157	98.1 %	2	1.3 %	160	100 %
Conversation strategy	9	10.1 %	9	10.1 %	71	79.8 %	89	100 %
Scaffolding	3	4.0 %	66	88.0 %	6	8.0 %	75	100 %
Triggering	8	13.3 %	29	48.3 %	23	38.3 %	60	100 %
Change of situation	21	87.5 %	1	4.2 %	2	8.3 %	24	100 %
Recipient	20	69.0 %	3	10.3 %	6	20.7 %	29	100 %
Word finding problems	0	0.0 %	18	100.0 %	0	0.0 %	18	100 %
Language economy/ lexical gap	0	0.0 %	14	93.3 %	1	6.7 %	15	100 %
We-code / they-code	0	0.0 %	5	100.0 %	0	0.0 %	5	100 %
Emotional affection	0	0.0 %	2	100.0 %	0	0.0 %	2	100 %
Indeterminate	0	0.0 %	10	83.3 %	2	16.7 %	12	100 %
total	70	6.0 %	949	80.9 %	154	13.1 %	1173	100 %

Table 2 Causes and Functions of CODEMIXING and CODEBREAKING (Multiple Entries Possible)

are described very briefly here. Havermeier (2015: 85-99) gives a comprehensive overview of the categories employed for the complete corpus analysis. Some of them have been intensively discussed in codeswitching research, e.g. linguistic domains or CODESWITCHING as a conversation strategy. Since the beginning of research on multilingualism, it has been described that different life domains are associated with different languages by most multilingual people and in most multilingual communities – a fact that causes CODESWITCHING when matters of different domains are discussed in one conversation (see, e.g., Fishman 1971). Related to this is the practice described by the terms *we-code* and *they-code*, which Gumperz (1982: 66) introduces. In this case, a certain language is used in an iconic way to signal affiliation with or dissociation from the group who typically uses it. Another topic of discussion in the literature has been the so-called *conversational codeswitching* (here called 'conversation strategy') that is described e.g. by Gumperz (1982), Auer (1984) and Gafaranga (2007). This practice is characterized by the fact that the change of language serves rhetorical purposes or conversation organization, regardless of the direction of the switch. Two further factors

that have ever since been described as crucial for language choice are the social situation of the conversation and the recipient (cf. Blom & Gumperz 1972: 422f). When those factors change, even the language used in a conversation might change, resulting in CODESWITCHING. The term *triggering* describes a psycholinguistic phenomenon, meaning the accidental choice of one or more words from another language than intended, triggered by preceding loanwords, cognates or intentional CODEMIXING (cf. Broersma & de Bot 2006). Further motivations for intentional CODESWITCHING can be *lexical gaps* or *language economy*, which means that one language provides a shorter expression than the other for something that the speaker wants to describe, with the consequence that the speaker chooses this expression, despite it not belonging to the base language. Even when speakers encounter difficulties finding the appropriate word, in a multilingual situation those may be solved by naming the corresponding expression in another language, or by using this language to explain what they mean to say. CODESWITCHING due to emotional affection has become a popular topic in the last decade and has been subject

to intensified research as well (see, e.g., Dewaele 2013).

The most common causes for CODESWITCHING in this corpus, *metalinguistic comments* and *reported speech*, as well as the so-called *scaffolding* are rather specific for the context the data has been compiled in. Scaffolding is a teaching practice, denoting repetitions in another language to ensure the recipient's understanding (cf. Kirkebæk 2013: 151f).

Furthermore, both in teaching situations and in academic discussions, it is quite common to quote others and to comment on expressions, definitions, etc. In a situation in which all interlocutors understand more than one language, quotes and reported speech can be produced in another language than the base language (most importantly in their original language, see also Example 4 in the previous section). Moreover, comments can be made in a different language, resulting in CODESWITCHING. This becomes visible e.g. in Example 6 given in the previous section, where the German unit under discussion "er ist krank" ('He is sick') is valuated in Swedish.

What is interesting about the meta-linguistic comments and reported speech in the corpus is not so much its strikingly high amount, but rather that the examined speaker community chooses to implement CODEMIXING in these occasions. When quoted or commented units are formulated in another language than the present base language, it would also be possible to use this language even for metalinguistic comments, resulting in CODEBREAKING. However, this opportunity is hardly ever taken. The same applies for, e.g., word finding problems and lexical gaps. When CODESWITCHING serves to solve these problems, this could be realized as CODEMIXING or cause a CODEBREAKING. Nevertheless, the speakers in this corpus never break with the base language in those occasions, but insert only the concerned expressions. This is especially interesting when it comes to language domains. In many examples given in the literature, the patterns would be classified as CODEBREAKING here, since it is described that domain-bound subjects would require a certain base language (see, e.g., Fishman 1971: 568; Clyne 2003: 159). This is not the case in my corpus, where language domains only become visible due to the insertion of a single expression, generally nouns, whose desig-

Example 6

A: wenn ich sage er IST krank↑ (2) *då har vi det glasklart↓ va↑* (1)
   when I   say he is   sick   then have we that crystalclear Q
   'If I say he is sick? (Pause) *Then we have it crystal clear, right*?

Example 7

A: das ist ja im   *sportlov↓*
   that is   PART in+the sportbreak
   'That is in the sport break (vacation in February).'

nated objects are associated with a specific country and consequently a specific language. An example of this is *sportlov* ('sport break') in Example 7. Since the speaker is working in the Swedish educational system, school vacations belong to the domain of Swedish for her. Moreover, the winter vacation in February, traditionally intended for winter sports, is a typical part of the Swedish academic year. In the German speaking countries, vacations at that time of the year are common only in few regions.

CODEBREAKING, on the other hand, is employed only when the function of CODESWITCHING can merely be fulfilled with a change of the base language. They are predominantly performed when the situation changes, or when a new recipient is addressed. In both cases, they can either be reactive or used to signal and establish the change.

This explains why the CODESWITCHING types contribute differently in different social situations. CODEBREAKING as a signal for or a reaction to a different addressee only occurs in situations in which the group of interlocutors is flexible. That is the case in lunchroom talks where side conversations with different interlocutors are common, in spontaneous student consultations with different students after a class, and in group work sequences during classes, where side conversations take place as well.

To sum up, some functions of CODESWITCHING are necessarily bound to specific structural outcomes, i.e. CODEMIXING and

CODEBREAKING. Others are not, but bilingual communities might establish regularities for them as well (see also Meyerhoff 2002). An analysis based on the suggested categorization can lead to interesting findings, especially regarding functions that may be realized through different forms of CODEMIXING.

5. Free and Bound Codemixing

5.1 Distinguishing Codemixing Based on Syntactic Dependencies

The term CODEMIXING, as defined here, subsumes a range of instances of bilingual speech. The embedded language can contribute with single words, such as in Example 3 and 8, with phrases such as in Example 9, as well as with clauses such as in Example 4.

For certain aspects of the analysis, it is necessary to distinguish between further subcategories. Besides the obvious option to distinguish between different word-classes, phrases, etc., one criterion that distinguishes the examples given above is to differentiate between whether or not the inserted elements are syntactically dependent on

base language units. The Swedish noun *mall* in Example 3, for instance, is a direct object in the German clause. That means that, inter alia, its case is dependent on the finite verb, *haben* ('have'), which demands two complements. The same applies to *avancerad nivå* in Example 9, which is dependent on the German preposition *auf*, demanding a complement in the dative. The discourse particle *va* in Example 8, on the other hand, is syntactically free and self-contained. The two Swedish clauses in the German conversation in Example 4 are, even though they are attached as a complement to *sagt* ('says'), not dependent on it in their form, and thus self-contained. I suggest calling the insertion of free and self-contained units FREE CODEMIXING. In contrast, incidences of CODEMIXING such as Example 3 and 9 are called BOUND CODEMIXING, modelled after the terminology of free and bound morphemes.

The category BOUND CODEMIXING comprises instances of CODESWITCHING that would belong to Muysken's category *insertion*, while FREE CODEMIXING contains instances that previous categorisations have parted or even sorted out. By highlighting the self-containment of the respective inserted units, FREE CODEMIXING also provides a

category for intersentential phenomena. Although some linguistic units, e.g. particles, are always instances of FREE CODEMIXING, the category is not restricted to certain word classes. All kinds of words and phrases can be used as FREE CODEMIXING when they are uttered outside or without a syntactical superstructure, for instance the Swedish noun in Example 10.

It can be assumed that a morphosyntactic dependency has certain implications for the formation and morphological integration of embedded elements.[5] Especially in inflectional languages, in which government and agreement markers are more important for the marking of syntactical functions than word order, the newly suggested categories are more suitable than Muysken's category *alternation*, which emphasizes the linear factor.

Evidence for this assumption comes from the German-Swedish corpus. Both of these languages are inflectional; along with that both have a relatively free order of clause constituents. The following analysis shows in which way the dependencies impinge.

---

5 Phonological integration may be affected as well, but phonological aspects were not part of the corpus study which this paper is based on.

5.2 Implementation of Free and Bound Codemixing in the Corpus

In the German-Swedish corpus, it was found that the CODEMIXING at the studied working places follows extremely strict patterns when it comes to its morphological and syntactical integration. More accurately, a morphological integration does almost never take place. Table 3 shows the results of an analysis of instances of CODEMIXING comprising one word belonging to a word class that may be inflected.

In the majority of the instances, there are no inflection morphemes to be found, thus it cannot be determined which language system is in charge. This does not mean that those instances are irregular bare forms, but that the required form is identical with the base form of the lexeme. This applies quite often, e.g. for nouns in the singular in most cases. However, if the embedded language words are inflected, this is performed with embedded language morphemes. Base language affixes are only used in very rare exceptions (0.5 % percent). Thus, a typical example of CODEMIXING including an inflected word is given in Example 11, where the Swedish noun *redovisning* is inserted in combination with a Swedish plural marker.

Example 3

```
A: also wir haben auch so ein mall↓
   PART   we  have  also  such a  template
```
'Well, we also have such a *template.*'

Example 4

```
(1)A: er sagt ja det är kvalitetshöjning↓ det gör vi↓
      he says PART that is  quality increase      that do we
```
'He says: *That is quality improvement. Let's do that*.

```
(2)B: hm↓ ja↓
       uhm ja
```
‚Um, ok.'

Example 8

```
A: da hatten wir das selbe genau↓ va↑
   there had   we  the  same  exactly Q
```
'There we had exactly the same, *right*?'

Example 9

```
A: ich halt es für SEHR wichtig dass wir also KURse auf dem
   I  consider it for very  important that we  PART courses on  the

   avancerad nivå haben↓
   advanced   level  have
```
'I think it is very important that we have, well, courses for the *advanced study period*.'

Example 10

```
A: Midsommardagen↓ Vielleicht hast du das gelesen↓
   Midsummerday-DEF.U.SG maybe       have you  that  read
```
'*The Midsummer Day*. Maybe you have read that.

```
B: hm ja
   uhm yes
```
'Uhm,yes.'

no inflectional morphemes	393	70.6 %
inflection according to base language	3	0.5 %
inflection according to embedded language	144	25.9 %
not determinable[1]	17	3.1 %
total	**557**	**100.0 %**

Table 3 Inflectional Morphemes in CODEMIXING Consisting of one Inflected Word
[1] due to breakoffs or homophone suffixes in both languages

Nevertheless, the base language actually does influence inflection with regard to reduction: There are some examples of words that are inserted without inflectional affixes in syntactical positions where the embedded language would demand inflection, but where the base language demands the base form. This can be seen in Example 12, where a Swedish adjective is used in predicative position in a German clause. In Swedish, predicative adjectives demand number congruence with the subject, so in this case, a plural marker would be needed. However, *lättklädd* is used without any plural marker, due to the German influence, because German demands predicative adjectives in their base form. The reduction only affects BOUND CODEMIXING, since only there we have a superior syntactic structure that is able to demand a certain form.

In a corpus in which words from the embedded language are not combined with base language affixes, regardless of their syntactical integration, how is it possible to tell that the distinction between FREE and BOUND CODEMIXING matters for morphological integration? In fact, it affects the definiteness marking in noun phrases. Definiteness marking is a question of inflection only in the case of Swedish definite noun phrases. In German (like in English), definiteness or indefiniteness are marked only by means of article words. Swedish has indefinite article words as well, but the regular definite article is a noun suffix. Also definite article words exist, but occur only as a complement to the definiteness suffix in certain syntactical circumstances (cf. Teleman et al. 1999a: 16f).

In the corpus analysis, the embedded language nouns have been examined with regard to whether they are combined with base language determiners, or with determiners from the embedded language. In the corpus examples hitherto given in this paper, there are already instances of different possible constructions. In Example 3 and Example 7, a base language article is allocated, whereas in Example 10, the noun is inserted together with its article. The latter can also be seen in Example 13 and Example 14.

Example 11

```
A:<lists> ÜBersetzung↓  ZUSAMMenfassung↓  (-)  schriftliche
           translation      summary                       written

   redovisning-ar↓  das  mach  ich  dann
   presentation-PL.U   that  make  I     then
```
'(lists) translation, summary, *presentations* of the writing abilities…I'm going to do that then.'

Example 12

```
A:  aber  die  sind  ja  nicht  lättklädd↓
    but   those  are  PART  not   scantilyclad-ø
```
'But they are not *scantily clad*.'

As Table 4 shows, the findings of the article analysis are not as clear as those concerning bound inflectional morphemes. Still, tendencies regarding the influence of BOUND and FREE CODEMIXING become visible. When embedded language nouns appear as FREE CODEMIXING, they are hardly ever combined with a base language article. Instead, they appear together with an article from the embedded language. In BOUND CODEMIXING, on the other hand, base language articles are preferred, but embedded language articles can also be found. Thus, the examples given above are quite representative for the findings in the corpus. The two examples in which bilingual noun phrases are built with a base language article and an embedded language noun both represent BOUND CODE-MIXING (Example 3 and 7). The instances including an embedded language article suffix are most often FREE CODEMIXING (Example 10 and 13), but this also happens with BOUND CODEMIXING (Example 14).

Whether or not embedded language articles are used, is also influenced by other factors than the question of BOUND or FREE CODEMIXING. One of these factors is proper names including an article, as is the case in Example 10. In those cases, the definiteness suffixes tend to occur in all cases of CODE-

Example 3
```
A: also  wir  haben  auch  so      ein     mall↓
   PART   we   have   also  such    a.ACC.N  template-ø
```
'Well, we also have such a *template.*'

Example 7
```
A: das  ist  ja    i-m      sportlov↓
   that is   PART  in+the.DAT.M  sportbreak-ø
```
'That is in the sport break (winter vacation in February).'

Example 10
```
A: Midsommardag-en↓ vielleicht hast du das gelesen↓
   Midsummerday-DEF.U.SG maybe       have you that  read
```
'*The Midsummer Day*. Maybe you have read that.

Example 13
```
A: das  ist  eigentlich das  was  ich= vägmärk-et↓
   that is   actually    that what I    trafficsign-DEF.N.SG
```
'That is actually what I… [meant]: *the traffic sign.*'

Example 14
```
A: dann hat aber= fakultärsnämd-en beschlossen
   then has but   facultyboard-DEF.U.SG    decided
```
'But then, *the faculty board* decided.'

MIXING. Another relevant factor for BOUND CODEMIXING is whether the superior structure is clausal or prepositional. In prepositional phrases, base language articles appear even more often than in verb complements. This is probably influenced by the fact that in German, articles in prepositional phrases often do not appear as a separate word, but as a clitic, as can be seen in Example 7 (*im* as the regular clitic form for *in dem*) (see also Nübling 2005: 107).

There are certain parallels between the findings on FREE and BOUND CODEMIXING in this corpus and the 4-M-Model, a supplement to the MLF-Model presented by Myers-Scotton and Jake (2000). This model does not only distinguish linguistic matter into *content morphemes* (those elected from the mental lexicon directly by the speakers' intentions) and *system morphemes* (roughly: functional morphemes and words, invoked for further formulation of syntactical structures together with implementation of content morphemes), but further distinguishes the latter into *early system morphemes* and *late system morphemes*, which then again are divided into *bridge late system morphemes* and *outsider late system morphemes*. Early system morphemes are directly invoked by certain lexemes, e.g. plural markers of nouns. Late system morphemes are invoked by a superordinate syntactic structure (cf. Myers-Scotton & Jake 2000: 3f). According to the MLF-model, the embedded language can provide content morphemes and early system morphemes in their company. Late system morphemes are less likely to be subject to CODESWITCHING (cf. Myers-Scotton 2002: 87f).

The findings in my corpus seem to support Myers-Scotton & Jake's assump-

All articles

	BOUND CODEMIXING		FREE CODEMIXING		total	
Base language article	59	72.0 %	2	4.4 %	61	48.0 %
Embedded language article	18	22.0 %	43	95.6 %	61	48.0 %
Both articles	5	6.1 %	0	0.0 %	5	3.9 %
total	82	100.0 %	45	100.0 %	127	100.0 %

Definite articles

	BOUND CODEMIXING		FREE CODEMIXING		total	
Base language article	42	65.6 %	1	2.7 %	43	42.6 %
Embedded language article	17	26.6 %	36	97.3 %	53	52.5 %
Both articles	5	7.8 %	0	0.0 %	5	5.0 %
total	64	100.0 %	37	100.0 %	101	100.0 %

Indefinite articles

	BOUND CODEMIXING		FREE CODEMIXING		total	
Base language article	17	94.4 %	1	12.5 %	18	69.2 %
Embedded language article	1	5.6 %	7	87.5 %	8	30.8 %
Both articles	0	0.0 %	0	0.0 %	0	0.0 %
total	18	100.0 %	8	100.0 %	26	100.0 %

Table 4
Distribution of Embedded Language Articles and Base Language Articles
in BOUND and FREE CODEMIXING of Nouns

tions: Most of the affixes that appear together with embedded language lexemes (and which all are embedded language morphemes) are early system morphemes. The articles, on the other hand, are late system morphemes, since they are needed to

include the nouns into a noun phrase. In those cases in which they indicate the function of the noun phrase as a specific verb complement, they are even outsider late system morphemes, since they are dependent on processes outside the phrase structure (cf. Myers-Scotton & Jake 2000: 4). Interestingly enough, exactly those morphemes are chosen from the base language,

in this case the matrix language. Moreover, they are more likely to be chosen from the base language when there is a superior structure demanding a case (BOUND CODE-MIXING) than when there is none, and consequently no matrix language (FREE CODE-MIXING).

Still, I do not claim that this is a universal pattern, but a pattern that can be found in

this very corpus, established as a routine in the speaker community. Evidence from other corpora shows that different speaker communities have very different preferences concerning definiteness marking of inserted nouns when mixing German and Northern Germanic languages (see, e.g., Kühl 2008: 113). What the evidence from my corpus does show is that there can be different patterns for the implementation and integration of BOUND and FREE CODEMIXING in a speaker community.

6. Conclusions

I have suggested a categorization of bilingual speech that in the first step distinguishes types of CODESWITCHING based on the main organization units of speech, i.e. turns and TCUs, and only in the second step is based on syntactic structures. The suggested categorization is summarized in Figure 1.

In this paper, it has been shown that instances of CODESWITCHING belonging to certain types according to this model share certain features not implied in their definitions. In my corpus, CODEMIXING and CODE-BREAKING are performed when the use of a second language has different functions and
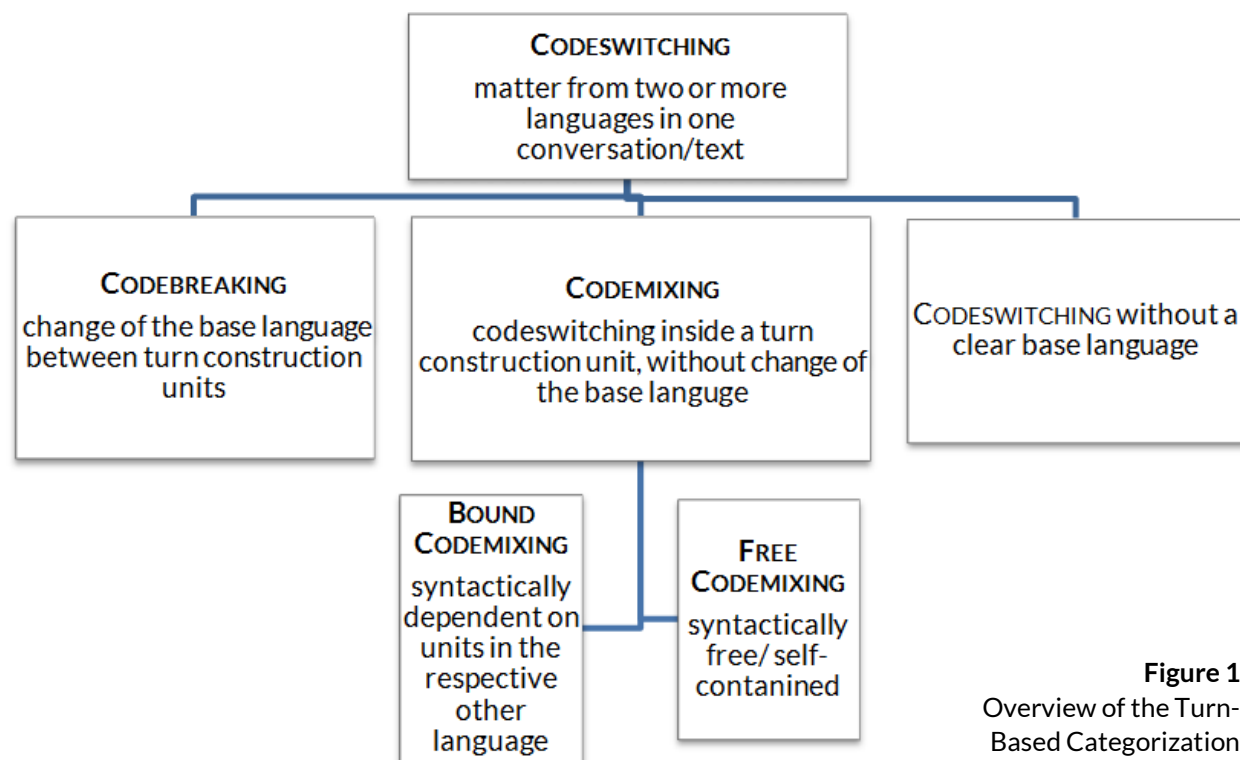

Figure 1
Overview of the Turn-Based Categorization of CODESWITCHING

motivations. Sometimes, the type of CODE-SWITCHING is necessarily determined by the function, but not always. It can be interesting to compare data from different speaker communities, especially for those functions that may be realized through different forms of CODEMIXING, to explore if they have established different routines and practices.

I do not claim that the patterns observed are universal, in the way that CODE-MIXING and CODEBREAKING always interact with the functions listed here. The same applies to the observations concerning FREE and BOUND CODEMIXING, which are treated differently in the analysed corpus when it comes to definiteness marking. As already mentioned above, different patterns concerning definiteness marking can be found in different speaker communities. What I suggest is that the types of CODESWITCHING that are defined here do have influence on the structures that can be found in bilingual speech. The outcome of the influence may be different depending on the language pair, the specific communicational needs of the observed speakers, etc. The suggested categorization provides a tool for further corpus analyses that can be useful for making patterns visible that were not considered in the categorization models used so far.

References

Auer, J. C. P. (1984). *Bilingual Conversation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Auer, P. (1999). From Codeswitching via Language Mixing to Fused Lects: Toward a Dynamic Typology of Bilingual Speech. *International Journal of Bilingualism*, 3.4, pp. 309-332.

Auer, P. (2009). On-Line Syntax: Thoughts on the Temporality of Spoken Language. *Language Sciences*, 31, pp. 1-13.

Backus, A. & Dorleijn, M. (2009). Loan Translations versus Code-Switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-Switching* (pp. 75-93). Cambridge: CUP.

Bhatt, R. M. (2013). Optimization in Bilingual Language Use. *Bilingualism: Language and Cognition*, 16.4, pp. 740-742.

Blom, J.-P. & Gumperz, J. J. (1972). Social Meaning in Linguistic Struture: Code-Switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Socioliguistics: The Ethnography of Communication* (pp. 407-434). New York et al.: Holt, Rinehart and Winston Inc.

Broersma, M. & de Bot, K. (2006). Triggered Codeswitching: A Corpus-Based Evaluation of the Original Triggering Hypothesis and a New Alternative. *Bilingualism: Language and Cognition*, 9, pp. 1-13.

Bullock, B. E. & Toribio, A. J. (Eds.) (2009). *The Cambridge Handbook of Linguistic Code-Switching*. Cambridge: CUP.

Clyne, M. (2003). *Dynamics of Language Contact: English and Immigrant Languages*. Cambridge: CUP.

Dewaele, J.-M. (2013). *Emotions in Multiple Languages*. Basingstoke: Palgrave Macmillan.

Dirim, İ. & Auer, P. (2004). *Türkisch sprechen nicht nur die Türken: Über die Unschärfebeziehung zwischen Sprache und Ethnie in Deutschland*. Berlin/New York: de Gruyter.

Duden. *Die Grammatik*. (2005) 7th edition. Mannheim et al: Dudenverlag.

Edel, K. (2007). *Strukturen des Bilingualismus – untersucht am Codeswitching Deutsch/Spanisch*. Frankfurt a.M./London: IKO Verlag für interkulturelle Kommunikation.

Fishman, J. A. (1971). Sociolinguistic Perspective on the Study of Bilingualism. In J. Fishman et al. (Eds.), *Bilingualism in the Barrio* (pp. 557-582). Bloomington: Indiana University Press.

Gafaranga, J. (2007). Code-Switching as a Conversational Strategy. In P. Auer & Li Wei (Eds.), *Handbook of Multilingualism and Multilingual Communication* (pp. 279-313). Berlin/New York: Mouton de Gruyter.

Gardner-Chloros, P. (2009). *Code-Switching.* Cambridge: CUP.

Grosjean, F. (1995). A Psycholinguistic Approach to Code-Switching: The Recognition of Guest Words by Bilinguals. In L. Milroy & P. Muysken (Eds.), *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching* (pp. 259-276). Cambridge: CUP.

Gumperz, J. J. (1982). *Discourse Strategies.* Cambridge: CUP.

Günthner, S. (2007). Ansätze zur Erforschung der „kommunikativen Praxis": Redewiedergabe in der Alltagskommunikation. In V. Ágel, & M. Hennig (Eds.), *Zugänge zur Grammatik der gesprochenen Sprache* (pp. 73-98). Berlin/New York: de Gruyter.

Havermeier, H. (2015). *Deutsch-schwedisches Codeswitching an der internationalen Universität.* Gothenburg: Department of Languages and Literatures, University of Gothenburg.

Kallmeyer, W. et al. (2002). *Variationsprofile: Zur Analyse der Variationspraxis bei den Powergirls.* Retrieved from http://www.ids-mannheim.de/prag/sprachvariation/fgvaria/Variationsprofile.pdf.

Kirkebæk, M. J. (2013). "Teacher! Why Do You Speak English?" A Discussion of Teacher Use of English in a Danish Language Class. In H. Haberland et al. (Eds.), *Language Alternation, Language Choice and Language Encounter in International Tertiary Education* (pp. 143-159). Heidelberg et al.: Springer.

Kühl, K. (2008): *Bilingualer Sprachgebrauch bei Jugendlichen im deutsch-dänischen Grenzland.* Hamburg: Kovač.

Leipzig Glossing Rules (2015). Retrieved from http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

Matras, Y. (2009). *Language Contact.* Cambridge: CUP.

Meyerhoff, M. (2002). Communities of Practice. In J. K. Chambers et al. (Eds.), *The Handbook of Language Variation and Change* (pp. 526-548). Malden (Mass.), Oxford: Blackwell.

Muysken, P. (2000). *Bilingual Speech: A Typology of Code-Mixing.* Cambridge: CUP.

Muysken, P. (2013). Language Contact Outcomes as the Result of Bilingual Optimization Strategies. *Bilingualism: Language and Cognition*, 16.4, pp. 709-730.

Myers-Scotton, C. & Jake, J. (2000). Testing the 4-M Model: An Introduction. *International Journal of Bilingualism*, 4.1, pp. 1-8.

Myers-Scotton, C. (1993). *Duelling Languages. Grammatical Structure in Codeswitching.* Oxford: Oxford University Press.

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes.* Oxford: OUP.

Nübling, D. (2005): Von *in die* über *in'n* und *ins* bis *im*: Die Klitisierung von Präposition und Artikel als "Grammatikalisierungsbaustelle". In T. Leuschner et al. (Eds.), *Grammatikalisierung im Deutschen* (pp. 105-131). Berlin/New York: De Gruyter.

Poplack, S. & Sankoff, D. (1988): Code-Switching. In U. Ammon et al. (Eds.): *Sociolinguistics. Soziolinguistik. An international Handbook of the Science of Language and Society. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft. Vol. 2* (pp. 1174-1180). Berlin/New York: de Gruyter.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en esnañol: Towards a Typology of Code-Switching. *Linguistics*, 18, pp. 581-618.

Sacks, H. et al. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50, pp. 696–735.

Sebba, M. (2012). Researching and Theorising Multilingual Texts. In M. Sebba et al. (Eds.), *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse* (pp. 1-26). New York/London: Routledge.

Sebba, M. et al. (Eds.) (2012). *Language Mixing and Code-Switching in Writing: Approaches to Mixed-Language Written Discourse.* New York/London: Routledge.

Selting, M. et al. (1998). Gesprächsanalytisches Transkriptionssystem. *Linguistische Berichte*, 173, pp. 91-122.

Selting, M. (2000). The Construction of Units in Conversational Talk. *Language in Society*, 29, pp. 477-517.

Söderlundh, H. (2012). Global Policies and Local Norms: Sociolinguistic Awareness and Language Choice at an International University. *International Journal of the Sociology of Language*, 216, pp. 87-109.

Teleman, U. et al. (1999). *Svenska Adademins grammatik. Vol. 3: Fraser*. Stockholm: Nordstedts Ordbok.

Treffers-Daller, J. (1997). Variability in Code-Switching Styles: Turkish – German Code-Switching Patterns. In R. Jacobson (Ed.), *Codeswitching Worldwide* (pp. 177-198). Berlin/New York: Mouton de Gruyter.

Appendix

I Transcription Conventions

normal font	German
italic font	Swedish
↓	falling intonation
↑	rising intonation
=	intonation break-off
(.)	falling intonation + pause shorter than 0.5 seconds
(-)	pause, shorter than 0.5 seconds
(1.5)	longer pause, length in seconds
< >	description of paraverbal activities, vocal color etc. (e.g. <laughs>)
majuscule	emphasized syllable
+	raised volume
°	lowered volume

II Abbreviations in the Glossing

ACC	accusative
DAT	dative
DEF	definite
IDF	indefinite pronoun
M	masculine
N	neuter
PART	modal particle
PL	plural
Q	question particle
REFL	reflexive pronoun
SG	singular
U	uter (common gender)