Bridges Over Troubled Waters: Theoretical Linguistics And Multilingualism Research[*]

Pieter Muysken

Centre for Language Studies, Radboud University, Postbus 9103, 6500 HD Nijmegen, Netherlands / Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre, Marais Street, Stellenbosch 7600, South Africa

E-mail: p.muysken@let.ru.nl

Abstract

This paper tries to construct a bridge between the concerns of theoretical linguistics and those of multilingualism and code-switching (CS) research. It argues that the primary special point of interaction between these fields lies in the question of potential equivalence between elements or categories, bridging across languages. After giving an overview of some major findings in recent CS research, these findings are interpreted in a constraint- or strategy-based framework. Then I explore the notion of categorical equivalence, starting with the observation that the insertion of single functional categories is highly restricted in CS contexts. Subsequently a number of concrete questions are formulated for research in this domain based on available data for Afrikaans-English and isiXhosa-English CS.

Keywords: code-switching, categorical equivalence, functional categories, isiXhosa, Afrikaans, English

1.   Introduction

This paper uses the bridge metaphor in two senses (hence the plural, with apologies to Simon and Garfunkel who released the legendary song with the singular bridge over forty years ago):

- It tries to construct a bridge between the concerns of theoretical linguistics and those of multilingualism and code-switching research; and
- it argues that the primary special point of interaction between these fields lies in the question of potential equivalence between elements or categories, bridging across languages.

The issue of categorical equivalence across languages has been hotly debated in recent publications. Is an adjective or a subject the same type of (basic or derived) category in different languages?

Haspelmath (2010:663) takes the stand of *categorical particularism*, citing Boas (1911:81), and states that descriptive categories chosen in accounts of particular languages "cannot be equated across languages because the criteria for category-assignment are different from language to language". Haspelmath goes on to argue that the "cross-linguistic comparison should be based on comparative concepts created by the typologist, rather than on crosslinguistic categories which are instantiated in different languages" (Haspelmath 2010:663).

After showing that the way different language descriptions use the term "dative" vary too much to define 'dative' as a universal category in the languages of the world, Haspelmath (2010:664) still goes on to argue that "*dative case* can be defined as a comparative concept". He then gives the following definition:

> A dative case is a morphological marker that has among its functions the coding of the recipient argument of a physical transfer verb (such as 'give', 'lend', 'sell', 'hand'), when this is coded differently from the theme argument. (Haspelmath 2010:664)

Comparative concepts in this perspective are thus (a) semantically based and (b) based on prototypical uses of a particular type of element. They function as useful heuristic devices to help typologists compare languages, but have no theoretical status as universal categories.

In contrast, Newmeyer (2010), reacting to Haspelmath and though not pleading directly for *categorical universalism*, the position taken in Chomsky's work, does advocate a more than purely operational characterization of comparative concepts. He argues that in actual practice, successful comparative concepts are more than purely heuristic devices, and should be rooted in real commonalities between languages.

Ultimately, however, Haspelmath and Newmeyer agree that in actual practice, linguists constantly look at categories in individual languages, but in the light of a more general perspective on properties shared by languages, even if the larger theoretical perspective adopted will differ. Some researchers (a typical example would be Baker (2003)) stress common and universal aspects of categories in the languages of the world whereas others, such as Haspelmath, take a particularistic perspective. The enduring question remains, of course, which aspects are specific to individual languages, and which are shared by many or all languages.

This, I want to argue, is where multilingualism and code-switching (henceforth "CS") research can make a contribution. Multilinguals manage and process several languages at the same time, and particularly when they code-switch, these languages interact. This interaction, in turn, frequently involves the mental computation, by speakers, of equivalences or the lack thereof. Studying CS thus offers a special window on cross-linguistic equivalence. Myers-Scotton (2006) phrases this as "natural codeswitching knocks on the laboratory door", although I will propose experimental work in this domain (rather than naturalistic observation) and in contrast with her work, a specific focus on cross-linguistic equivalence.

It should be noted that CS research is directly informed by and contributes to theoretical linguistics since it concerns language structure and its analysis in bilingual utterances. However, the special and unique contribution of CS, in my view, comes from the study of cross-linguistic equivalence.

The argument is built up as follows. In section 2 I give an overview of some of the major findings in recent CS research, and these findings are interpreted in section 3 in a constraint- or strategy-based framework. In section 4 I explore the notion of categorical equivalence, starting with the observation that the insertion of single functional categories is highly restricted in CS contexts, and in section 5 a number of concrete research questions are formulated to study equivalence. Section 6 concludes this paper.

2.   Major findings in CS research

In recent years a large number of studies have been done on CS, the alternate use of two languages within the same speech event. These studies cover sociolinguistic, pragmatic, psycholinguistic, and grammatical domains (cf. e.g. Muysken 2011). In the grammatical domain, the main focus has been on the clause: how can fragments of two languages be combined in a clause in such a way that the overall result is grammatical?

The main methodology so far has been the analysis of recorded naturalistic CS data, although recent studies also involve experimental techniques (see further in section 5). Meta-analysis of the findings up to 2000 leads to the following general conclusions (Muysken 2000).

- There is maximal integrity of different participating language fragments in CS.
- Linear equivalence between the languages favours CS.
- The peripheral position of the switched fragment in the sentence structure favours CS.
- There are restrictions on the insertion of single functional elements in CS.

I will illustrate these findings with data from Sinhala-English CS presented and analyzed by Senaratne (2009). The phenomenon of maximal integrity of different participating language fragments in CS is illustrated in (1). The first part of (1) is well-formed English, the second part well-formed Sinhala:

(1)   *I will tell him again*   giyee   naet-nan
                                  go.EMP   NEG-CMP
       "I will tell him again, if he does not go."   (Senaratne 2009:2)

The linear equivalence between the languages involved in CS is illustrated in (2) and (3). In (2) there actually is 'doubling' between the English preposition *from* (creating an English-style P DP order) and the Sinhala ablative case marker *–in*:

(2)   *He is a southerner from*   Bentɛrɛ   gang-in   ehaa
                                    Bentara   river-ABL beyond
       "He is a southerner from beyond Bentara river."   (Senaratne 2009:208)

In (3) the Sinhala adverbial clause precedes an English main clause:

(3) ma-Tɛ kiyannee naetuvɛ   *she went ahead and wrote the letter*
     1sg-DA say.EMP   NEG.PAR.RL
     "Without saying to me she went ahead and wrote the letter."   (Senaratne 2009:209)

That the peripheral position of the switched fragment in the sentence structure favours CS is illustrated in (4) and (5). Again, in (4) we have a Sinhala adverbial clause – English main clause sequence:

(4) oyaa *horror films* balɛnɛ-koTɛ   *I don't criticize that no?*
     2sg   horror films   watch.RL-CMP
     "When you watch horror films, I don't criticize that, do I?"   (Senaratne 2009:217)

In (5) we have a loosely coordinated structure beginning with an English main clause followed by the Sinhala emphatic marker *ko*, and then a coordinate Sinhala clause:

(5) *Let him come*   ko,   mamɛ dennan   eyaa-Tɛ
                       EMP   1sg   give.VL   3sg-DA
     "Let him come will you, then I will deal with him."   (Senaratne 2009:213)

The final major finding, that is, that there are restrictions on the insertion of single functional elements in CS, is also amply illustrated by the Sinhala-English data, particularly in a negative sense: such elements are largely lacking in the data. The major patterns in the Sinhala-English corpus, as indeed in other corpora as well, are illustrated in Table 1, showing a majority of content categories inserted:

Table 1. English lexical categories and phrases in Sinhala utterances

#	Categories
386	English nouns/noun phrases in Sinhala utterances
58	English adverbs/adverbial phrases in Sinhala utterances
69	English verbs and verb phrases in Sinhala utterances

Some adverbials do occur, such as the Sinhala English expression *first shy* in (6):

(6) saamaanyen *insurance* ek-ak   *first shy*-mɛ dennɛ   oona
     usually   insurance NM-IND   first shy-EMP   give.INF   should
     "Usually insurance is given on the very first shy."   (Senaratne 2009:165)

Potentially more problematic for a generalized ban on the insertion of single functional elements in CS is the apparent presence of inflected verbs, as in (7)-(10). In (7) and (8) an apparent verb occurs in past participle form, but in (7) it is really an adjective and probably it is treated as such as well in (8), where there is no further auxiliary present to trigger a past participle interpretation:

(7) eyaa harimɛ *worried* nee?
     3sg   very   worried   EMP
     "He is very worried isn't he?"

(8)   *program*   ɛkɛ   *changed*-lu?
       program   NM.DF   changed-EMP
       "The program has changed it seems."   (Senaratne 2009:173)

In (9) and (10) we have English verbs with a progressive participle ending. Again, however, there is no element in the morphosyntactic environment triggering this progressive ending, and these cases may be best viewed as frozen forms with an interpretation of ongoing or immediate future action:

(9)   mokak dɛ anee   oyaa   taamat   *writing* dɛ?
       what   Q   INT   you   still   writing   Q
       "Are you still writing?"

(10)   oyaa   *coming*   nee   da?
       you   coming   EMP   Q
       "You are coming aren´t you?"   (Senaratne 2009:173)

Another set of cases concerns the frozen forms *pass* and *fail*, which occur without any inflection, either in Sinhala or English.

(11)   panti-yee   lamay   okkoomɛ   *fail*?
       class-GEN   child.PL   all   fail
       "All the children in my class have failed?"

(12)   oyaa-gee   *class*   ekee   innɛ   lamay   okkoomɛ   *pass*
       2sg.GEN   class   NM-GEN   be.RL   child.pl   all   pass
       "All the children in your class have passed."   (Senaratne 2009:169)

A final case concerns the occurrence of the single negation element *no*. There are two occurrences in the corpus collected by Senaratne:

(13)   *No* kata *no* sina   [fixed expression]
       no talk no laugh
       "If there is no talk then there is no laughter." (Senaratne 2009:176)

(14)   eyaa-Tɛ   *no*   sellam   maa-t   ekkɛ   [pidgin-like]
       3sg-DA   not   game.PL   1sg-also   with
       "He will not be able to play with me."   (Senaratne 2009:177)

The first one (13) is a fixed expression, and in the second one (14) it is a pidgin-like construction. Notice that the clausal negator is *no* rather than *not* here.

For the rest, there are no single elements from English in the Sinhala utterances: no pronouns, articles, prepositions, complementizers, etc.

3.   A first approach to the constraints on CS

The approach I have taken in my own work on CS (Muysken 2000) is to create a typology of the relevant phenomena. In my view this typology is observationally still fairly adequate and

does cover a large part of the phenomena in CS. It distinguished between *insertion* of elements into a large structure, *alternation* between fragments from different languages, and *congruent lexicalization* of different words in patterns largely shared by the two languages.

The big problem with this approach is that it does not naturally account for the factors and constraints relevant to particular choices in CS (processing, social, competence, etc.). There is also no single unified speaker model in this approach, only a taxonomy, and hence no explanatory adequacy.

A possible solution to this is to work with competing speaker strategies, comparable to Optimality Theoretic "constraints" (Prince and Smolensky 2004). Different rankings of these strategies would then produce different switching outcomes, such as the three CS types proposed in Muysken (2000). A first such strategy would be Select.

SELECT
Use elements from the language most suited …
      to express a specific cultural content,
      to express a particular relationship between interlocutors,
      in a particular setting,
      …

Values for Select could be: Select L1, Select L2, Neither specifically, Both.

Notice that Select is not specific to CS. Rather it is a general strategy language users need to follow in their daily lives. Of course, it is particularly brought into relief in bilingual contexts, since bilingual speakers have to select constantly. A second general strategy is Contour.

CONTOUR
Create contour, relief or contrasts in your message:
      through language selection
      in pauses
      intonation
      word choices
      …

Contour is what separates interesting speakers from uninteresting ones, effective conversationalists from ineffective ones. It allows us to use language strategically in order to organize our discourse. It is thus a general, but not obligatory strategy. Contour is not limited to bilingual contexts, but may involve different languages. An equally general strategy is Max Weight:

MAX WEIGHT
A fragment in a particular language should contain at least one stress contour or independent tonal contour, but preferably as many as possible.

Max Weight is a general well-formedness constraint on utterances, guaranteeing that they are sufficiently long to be easily pronounceable following the rules of the language involved, but also leading to a preference for longer fragments in the same language within bilingual utterances. An example of the effect of Max Weight can be seen in the following contrast in

approval in a magnitude estimation task in Van Dulm (2007). Afrikaans-English bilingual students were asked to comparatively rate different bilingual utterances, and (15a), containing a longer English fragment, was ranked significantly better than (15b) with a shorter fragment:

(15)   a.   Ons ouers   dink dat daardie groot voëls *catch them in flight*. 1.18
              our parents think that those   big   birds

        b.   Ons ouers   dink dat daardie groot voëls *catch them*. 1.01
              our parents think that those   big   birds

Again Max Weight is operant in CS, but it is in fact a more general principle holding for utterances in general. The same holds for F-Agree:

F-AGREE
Comply with all language-specific feature checking requirements in a specific structural domain:
    person, number, case, agreement, …
    subcategorization features
    structure-based interpretation

The same magnitude estimation task in Van Dulm (2007) shows the power of this constraint:

(16)   a.   Ons ouers dink dat daardie groot voëls *catch them in flight*.   1.18
              our parents think that those   big   birds

        b.   Ons ouers   dink dat daardie groot voëls *them catch in flight*.   .79
              our parents think that those   big   birds

Cases in which the correct configuration (from the perspective of English) holds for the elements 'catch' and 'them' are evaluated much better than the alternative. In any case, this strategy (and it has been given many formulations) is really the backbone of grammaticality in language, and hence by no means exclusive to CS, although it holds in a powerful way in many CS utterances as well.

Another important strategy is Combine, which simply concerns the possibility of combining elements in language:

COMBINE
Combine elements in an utterance, leading to …
(a) combinations in feature checking, i.e. F-Agree configurations
(b) combinations in adjunction configurations
    with pauses
    adverbial elements
    extraposed elements
    …

In Combine it is possible to either combine elements in F-Agree configurations or randomly, leading to paratactic or extra-grammatical utterances. The reason I draw attention to this possibility is because it is hard to classify some bilingual utterances as grammatical from the

perspective of either of the languages concerned; rather, they seem to reflect some kind of juxtaposition.

F-LINK
Establish correspondences between categories and features in the two languages involved in a CS utterance

---

This 'strategy' or 'constraint' is specific to bilingual settings, in contrast to the other constraints. Although the spirit of the approach taken in this paper is minimalist in a broad sense, I emphatically want to contrast the approach taken here to the work of e.g. McSwann (1999), which assumes that we can dispense with something like F-Link because features and categories are universally defined.

I will return to F-Link extensively below, and want to close this section by discussing the issue of constraint rankings. A first set of rankings concerns the conditions for CS in the first place:

CONTOUR   MAX WEIGHT
MAX WEIGHT >   switch   CONTOUR   >   non-switch

If Max Weight is ranked above Contour, there is no switch in the first place.

A second set of rankings concerns two types of configurations. I will assume that paratactic configurations are like adjunction, and syntactic configurations involve checking:

COMBINE   F-AGREE
F-AGREE   > adjunction configuration   COMBINE   > checking configuration

A third set of rankings involves the placement of F-Link. If it is ranked high, 'proper' CS is likely, satisfying equivalence between languages, and if it is ranked low, often semi-licit switches result:

F-LINK   ….
….   > "proper" CS   F-LINK   > "improper" CS

A provisional set of rankings for the three types of CS discussed in Muysken (2000) may be:

Insertion	*Alternation*	*Congruent lexicalization*
CONTOUR	CONTOUR	CONTOUR
MAX WEIGHT	MAX WEIGHT	SELECT L1&L2
SELECT L1&L2	SELECT L1&L2	MAX WEIGHT
F-AGREE	COMBINE	F-AGREE
F-LINK	F-AGREE	F-LINK
COMBINE	F-LINK	COMBINE

The reader should bear in mind that these different strategies, and hence these different rankings, may also be linked to different sociolinguistic considerations.

4.   The special status of functional categories and the issue of equivalence

In the discussion of Sinhala-English CS in section 2 it was noted that the English elements inserted into Sinhala clauses were limited to a few types. In particular, it is evident that functional elements cannot be inserted by themselves. This result, one of the most significant findings in the CS literature, has been explained by Myers-Scotton (1993) in terms of processing, in particular the multilevel processing model of Garrett (1975, 1980) and related models. However, this explanation cannot be fitted very well into many recent accounts of language processing.

4.1.   Typological equivalence and grammaticalization

In this section I want to propose an alternative explanation: the hierarchy of elements that can be inserted by themselves in CS (high insertability: nouns; low insertability: agreement markers) matches the typological equivalence hierarchy (distinguishable in most or all languages: nouns; highly variable across languages: agreement markers) as discovered by typologists almost exactly. This is presented in Table 2.

Table 2. Matching two hierarchies: the content-system hierarchy in CS research, and the typological equivalence hierarchy

Content-system hierarchy	*Typological equivalence hierarchy*
Noun	Noun-verb distinct in most languages
Adjective	Adjectives small class in many languages
Verb	Adjectives ~ nouns in some languages: ~ verbs in some languages
Adverb	Adverbs fairly general but undefined lexically
Adposition	Adpositions may be absent and vary in richness
Coordinating conjunction	Coordinating conjunctions often present
Subordinating conjunction	Subordinating conjunctions vary in form in different languages
Pronoun	Pronouns often present but very diverse feature systems
Determiner	Determiners may or may not be present
Case marker	Case realized or not various ways
Agreement marker	Agreement varies widely across languages

This explanation has two additional advantages over the one in Myers-Scotton (1993):

(a)   It is independently needed, since it makes it possible to take structural language distance into account: the more similar two languages are, the easier it is to insert elements lower on this hierarchy;

(b)   It allows for a fine-grained hierarchy; the approach taken in Myers-Scotton and Jake (2000) does imply a four way division, but this division is not well-rooted in the psycholinguistic data they adduce.

However, the approach proposed here crucially relies on the notion F-LINK: only if we assume F-LINK does this hierarchy of elements like nouns at the one end and agreement markers at the other come into play.

One of the main sources for our conception of cross-linguistic equivalence of grammatical categories comes from the grammaticalization literature (Hopper and Traugott 2003) in

historical linguistics. In principle, for a matrix language (ML) and an embedded language (EL) in CS, there are five possibilities with respect to the grammaticalization of grammatical features, as shown in Table 3.

Table 3. Possibilities for feature matching in grammaticalization patterns

	Matrix language	*Embedded language*	*Features*	*Examples*
A	X	X	Grammaticalized the same way in ML and EL	fem-masc in French and Spanish
B	X	-	Grammaticalized in ML but not EL	English nouns in Spanish
C	-	X	Grammaticalized not in ML but only in EL	Spanish nouns in English
D	X	Y	Grammaticalized differently in both languages	3 gender German, 2 gender French
E	-	-	Not grammaticalized in either language	Evidentiality in French and English

4.2.   Dimensions of equivalence: psychotypology

Typological similarity can be defined in several different ways. One way would be genealogical relatedness: languages share many features if they are daughters of the same language. However, languages also start diverging from their ancestor language, leading to typological contrasts. A simple instance is the word order contrast between English on the one hand, and in fact closely related languages such as Afrikaans and Dutch on the other.

A second way would be typological comparative work, often involving the study of alternative ways or expressions of a single notion or proposition in different languages. This type of work is a key component of the research programme defended here, but it has the potential disadvantage that it is not known ahead of time which cross-linguistic differences or similarities are important to bilingual speakers and which ones are not.

For this reason, a third option can be thought of, in terms of psychotypology as it was defined by Kellerman 1979). The notion of psychotypology explicitly distinguishes 'objective' structural similarity of specific linguistic features from subjectively perceived similarity. This has the great advantage of taking language distance out of a mechanistic sphere in which similarity is externally defined and into the sphere of cognition. However, it also makes it an intractable notion and hence a potential risk for circular reasoning (of the type: this switch does not occur and must therefore reveal incompatibility in the mind of the speakers).

To reduce the risk of circularity, it is possible to take recourse to the strategies proposed by Sebba (2009), which mark the specific ways multilingual speakers establish equivalence between categories:

---

harmonization: match existing features
neutralization: add "masking" morphological material (Lq carriers)
null: omit offending category in one language
compromise between feature systems

Harmonization is illustrated in (17) and (18). In (17) the German passive auxiliary *wurd* would select a German passive participle; however, what occurs is the French passive participle *recalé* "failed".

(17)   Wann der client *recalé* **wurd** am permis.
        when the client failed   is   at licence
        "When the client fails the driving test."   (Gardner-Chloros 1991)

In (18) the Spanish progressive auxiliary *está* would select a Spanish present participle ending in *–ndo*; however, what occurs is an English present participle ending in *–ing*.

(18)   Siempre **está** *promising* cosas
        "Always (she) is promising things."   (Poplack 1980)

These equivalences are not remarkable in themselves. After all, the languages are related and the constructions similar. Nonetheless, they are not automatic and not automatically provided by a theoretical construct. Many languages lack participles, past or present. Moreover, even some languages closely related to those involved in (17) and (18) lack these exact constructions; that is, they have not been grammaticalized in such languages.

Sometimes F-LINK may be the result of processes of convergence over time. An example cited by Sebba (2009) comes from the study of Swahili-English CS. In this example the copula is recruited to form a passive construction together with the English past participle. In reality, the passive is formed in Swahili with a passive suffix on the verb, as in (20).

(19)   **I-li-ku-wa**   *discussed* kwenye *approximants*
        CL9-PAST-INF-BE   discussed under   approximants
        "It was discussed under approximants."   (Kibogoya 1995)

(20)   wa-li-pig-**wa**
        3PL-PAST-beat-PASS
        "They were beaten."

5.   A research programme 1: Afrikaans-English CS

In Ingrid Winterbach's otherwise rather serious Afrikaans novel *Die benedenryk* (2010), there is a comical interlude in which a young artist is introduced, Jimmy Harris (probably from the southern suburbs of Cape Town, the protagonist surmises). This young artist, as many of his peers are wont to do, rants against the current art scene in South Africa:

(21)   … Dis nog nie *multimedia* genoeg nie. Dis nog nie *take no prisoners* nie. Dis nog nie
        confrontational genoeg nie. Dit *challenge* nog nie sy eie aannames *stringently* genoeg
        nie. Dis nog nie genoeg van 'n *assault* op enige *established high culture* nie.

"... It's not multimedia enough yet. It's not take no prisoners yet. It's not yet confrontational enough. It doesn't challenge his own assumptions stringently enough yet. It's not yet enough of an assault on any established high culture."

(Winterbach 2010:61)

It is not clear, to me at least, whether the novelist is portraying a CS monologue or an Afrikaans version of an English monologue. In any case, the cases of CS portrayed stem from the author's ear for usage or creative imagination, but fragments like these raise a number of issues.

We find single occurrences of English nouns, adjectives, verbs, adverbs, as well as full constituents and expressions from English. Are these constituents fully interchangeable with Afrikaans elements of the same type, suggesting full categorical equivalence? Based in part on Sebba (2009), we can distinguish four possibilities (set out in (A) to (D) below):

(A)   simple *harmonization*: match existing features, as in many of the insertions mentioned above;

(B)   *neutralization*: add 'masking' morphological material. An example in Winterbach's material concerns the prefix *ge-* on English verbs. There are a number of cases of this with passive verbs:

(22)   a.   alles   kan ge-*justify*   word
             everything can PAST-justify become
             "everything can be justified"   (Winterbach 2010:63)

        b.   dit word ge-*rule*   deur die *tyranny of* die *object*
             it is   PAST-rule by   the tyranny of the object
             "It is ruled by the tyranny of the object"   (Winterbach 2010:64)

        c.   Presies   waar dit ge-*fuck*   moet word
             precisely where it PAST-fuck must be
             "Precisely where it must be fucked"   (Winterbach 2010:64-5)

However, with past tense *ge-* there is some variation:

(23)   a.   die *boundaries* het   *collapse*
             the boundaries have collapse
             "the boundaries have collapsed"   (Winterbach 2010:65)

        b.   iemand   soos X ... en   die *effects* daarvan ge-*document*   het
             someone like X ... and the effects thereof PAST-document has
             "someone like X ... and the effects thereof have documented"
                                                               (Winterbach 2010:65)

Three questions for further corpus studies and experimental work come to the fore right away: (i) Is there indeed a difference between the realization of the passive *ge-* and past tense *ge-* prefixes? (ii) Which factors (lexical, phonological, morphological, semantic, frequency-

related) govern the presence / absence of *ge-*? (iii) Are there any differences between speakers in this respect?

In the Cape Flats radio talk show corpus transcribed by Bowers (2006), the majority of inserted English verbs, either in past tense (24)-(26) or passive (27)-(29) form, are marked with *ge-*. (The numbers after the examples refer to the turns of the transcript attached to the thesis; see also the Appendix for the other cases in the transcript by Bowers 2006.)

Past tense:
(24)   *No … I was the navigator of this tour* en ek het ge-*navigate*, verstaan jy
                                    "and I navigated, you understand"   (Bowers 14)

(25)   Hy het nie eens ge-*argue* nie
        "He didn't even argue"   (Bowers 57)

(26)   Die floral dress, die floral curtains het ge-*clash* met alles, verstaan jy
        "The floral dress, the floral curtain clashed with everything, you understand"
                                                                       (Bowers 623)
Passive:
(27)   Mense word ge-*dump*. En dan *argue* hulle nou wie't vir wie ge-*dump*, verstaan jy
        "People get dumped. And then they argue about who dumped who, you understand"
                                                                       (Bowers 182)

(28)   en al wat die mense wil gehad het is, hulle wil hulle hare ge-*blowdry* het
        "and all the people wanted is, they wanted to blowdry their hair"   (Bowers 343)

(29)   die assets word ge-*share* vanaand, verstaan jy
        "the assets will be shared tonight, you understand"   (Bowers 532)

Only in one case is *ge-* absent, and this involves a CS across two turns:

(30)   S1:   Uh, um, is daar wat hy sy, sy, sy *information*
                "it's there that he got his, his, his information"
        S2:   *theories develop* het
                "developed his theories"   (Bowers 408/9)

In another case, there was no *ge-*, but the past participle was part of an English "island" (Myers Scotton 1993), as shown by the English adverb. Notice also the presence of English *-ed*:

(31)   *because* ek is nou *officially promoted*
                "I am now"   (Bowers 10)

In yet another example, there is no *ge-*, but in that case the past participle is clearly adjectival, and also marked with *-ed*:

(32)   *but*, hy, hy't gesê sy pa gat uitwerk daar's'ie 'n *joke* hier *involved* nie
        "but he, he said his father was going to work out that there isn't a joke involved here"
                                                                       (Bowers 129)

This would lead to an additional research question: (iv) are *ge-* and *–ed* ever compatible for speakers?

 A final comment is that sometimes speakers may be hesitant about prefixing *ge-*:

(33)   toe, toe (XX) die hele *shed* uitmekaar uit, *you know*, ge- ,ge-*horticulture* het, verstaan jy
         "so, so (XX) the whole shed apart, you know, he horticultured it, you understand"
                                                                          (Bowers 738)

Notice that *ge-* can also be attached to words from Arabic:

(34)   Yes, I'm well NAME. Kla, klaar ge-*jummuah* en alles nou
                              "Finished with jummuah (Islamic prayer) and all now"
                                                                          (Bowers 911)

A second phenomenon that could be studied, illustrated by Winterbach's character, is the use of the infinitive particles *om* and *te* after verbs such as *probeer*:

(35)   om dit te *reclaim* van   die *heteronormative structures* wat   dit probeer *naturalize*
         INF it to   from the   which it try
         "to reclaim it from the heteronormative structures that that try to naturalize it"
                                                                          (Winterbach 2010:66)

Under certain circumstances *te* can appear in these constructions, as in the following example from a Facebook exchange:

(36)   Hy probeer Afrikaans vorentoe **te** bring, en die is goed.
         "He tries to bring Afrikaans to the fore, and that is goed."

In the Bowers corpus, there are several cases where the verb *probeer* is replaced by *try*, and then both *om* and *te* appear:

(37)   *Try* jy **om** vir my, *try* jy **om** daai van my weg *te* vat en ek slat jou in jou *you know where*.
         "Try to take that away from me and I'll hit you in your you know where" (Bowers 207)

(38)   Is, is, Clarence gat (gaan) *try* **om** so … funky **te** hou, verstaan jy.
         "It's, it's, Clarence is going to try to keep himself funky, you understand."
                                                                          (Bowers 881)

Research questions here would concern the absence and presence of *om* and *te* when either the matrix or the embedded verb is taken from the other language.

(C)   ***null***: omit offending category in one language, generally the inserted one. This is of course related to the null strategy, of omitting (in this case morphologically) offending elements. In example (39) (from the Bowers corpus) the utterance starts in English and switches to Afrikaans. However, the English verb *phone* occurs, but without the preverbal auxiliary *het* or the prefix *ge-*:

(39)   *I was on my way to, uh, an appointment in Sea Point* en toe *phone* die *client* om vir
                                                                    "and so this client phoned to
          my te sê … Toe *phone* ek nou so vir Clarence.
          tell me … So I phoned Clarence"   (Bowers 566)

Neither is the appropriate English tense marker *–ed* present, of course. What is the incidence of such null forms?

(D)   *compromise* between the feature systems of the two languages. Here it is possible that e.g. resemblances between Afrikaans *te* and English *to* play a role, to return to the early set of *probeer* examples.

The complexity of the phenomena involved, the variability in the findings up to now, and the interaction of different intervening factors all call for an experimental approach with structured elicitation in addition to the corpus-based research that has been carried out so far. Gullberg et al. (2009) provide a more reasoned outline of this, sketch various possibilities, and point to the growing literature exploring this kind of approach.

6.   Research programme 2: isiXhosa-English CS

Much less is known about the interaction of Afrikaans and English with the other languages of South Africa. Starting with pioneering studies such as Janson (1983), Gilbert and Makhudu (1984), and Schuring (1985) and related work from the same period, there has been a tradition of studies on mixed urban contact vernaculars, but less on spontaneous CS as such, in the absence of a stable mixed vernacular.

While Simango (2007) has worked on the grammatical consequences of inserting English verbs into isiXhosa utterances, the main issue that has been addressed in the literature on isiXhosa-English CS is the question of integration of English nouns into isiXhosa with isiXhosa noun class prefixes. The principal study is De Klerk (2006). She concludes that the *i-* prefix is by far the most common one, and that with humans *u-* and *ama-* are preferred. Other elements used are *isi-* and *aba-*. It is not clear exactly which factors are involved in noun class assignation, in addition to semantic ones. Allwood et al. (2010:88) provide a list of the ten most frequent English words with isiXhosa prefixes in the UNISA corpus of spoken languages of South Africa. Interestingly enough, not all words have the same prefix as in the data provided by De Klerk (2006).

(40)   ii-drugs   (compare *i-drugs* in De Klerk 2006)
          i-aids
          i-crime
          i-right
          eyi-one
          i-hiv
          e-town
          i-government   (compare *u-government* in De Klerk 2006)
          ii-firms
          i-chance

English words integrated phonologically into isiXhosa also receive prefixes (e.g. Kirsch and Skorge 2001:35-36):

(41)   i-filimu   "film"
        i-folokhwe   "fork"
        isi-tovu   "stove"
        u-titshala   "teacher"

However, occasionally a word appears without a prefix (e.g. Kirsch and Skorge 2001:34):

(42)   Le **rowuzi** ibomvu krwe/krwee.
        "This rose is blood red."

Presumably there are also phonological reasons for not prefixing certain words, but this needs further investigation. It is also clear (e.g. Mati 2003) that there is strong normative pressure for some speakers against using English or Afrikaans words or phrases in isiXhosa, but again no systematic research has been carried out on this.

Again, the corpus-based approaches used so far could and should be complemented with experimental approaches in which semantic, morphological, phonological and frequency factors are systematically explored to see how English words are incorporated into isiXhosa, and what the factors are that condition their acceptance. Once more is known about other aspects of isiXhosa-English CS, these experiments could be extended into those domains as well.

7.   Conclusions

In this paper I have
- sketched the theoretical issue of categorical particularism versus universalism and its potential relevance for CS studies;
- given an all too brief perspective on the current state of knowledge in this area;
- provided a glimpse of the possibility of a constraint-based approach to CS studies and in particular of research on categorical equivalence and congruence; and
- briefly illustrated possibilities for research in this domain for Afrikaans-English and isiXhosa-English CS.

The main point of all this was to hint at the exciting possibilities for cross-pollination between theoretical linguistics and multilingualism and CS research, and at the need for experimental work in this domain.

Abbreviations used in glosses

ABL	Ablative	DA	Dative	IND	Indefinite	NM	Nominalizer	PL	Plural
BE	Copula be	DF	Definite	INF	Infinitive	PAR	Past participle	Q	Question marker
CL	Classifier	EMP	Emphatic	INT	Interrogative	PASS	Passive	RL	Relative marker
CMP	Complementizer	GEN	Genitive	NEG	Negative	PAST	Past tense	VL	Volitive

References

Allwood, J., H. Hammarström, A. Hendrikse, M.N. Ngcobo, N. Nomdebevana, L. Pretorius, and M. van der Merwe. 2010. Work on Spoken (Multimodal) Language Corpora in South Africa. In N. Calzolari et al. (ed.) *Proceedings of LREC 2010*. Malta, European Language Resources Association (ELRA), 885-889.

Baker, M.C. 2003. Lexical Categories. Verbs, Nouns and Adjectives. Cambridge: Cambridge University Press.

Boas, F. 1911. *Handbook of American Indian Languages*. Washington, D.C.: Smithsonian Institution. Bureau of American Ethnology.

Bowers, D.L. 2006. *Grammatical Constraints And Motivations For English/Afrikaans Codeswitching: Evidence From A Local Radio Talk Show*. MA Thesis, Department of Linguistics, University of the Western Cape.

Dulm, O. van. 2007. *The grammar of English-Afrikaans code switching. A feature checking account*. Doctoral dissertation, Radboud University, Nijmegen LOT-series.

Gardner-Chloros, P. 1991. *Language Selection and Switching in Strasbourg.* Oxford: Clarendon Press.

Garrett, M.F. 1975. The analysis of sentence production. In G. Bower (ed.) *The psychology of learning and motivation*, vol. 9. New York: Academic Press, 133-177. Reprinted in Gerry T.M. Altmann (ed.) (2002) *Psychology and Linguistics. Critical concepts in psycholinguistics V*. London: Routledge, 34-75.

Garrett, M.F. 1980. Independent processing levels in sentence production. In V. Fromkin (ed.) *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press, 263-71.

Gilbert, G. and D.P. Makhudu. 1984a. Ian Hancock's 50 sentences (for comparative study of the English-based Atlantic creoles) translated into Standard European Afrikaans (SEA), Coloured Afrikaans (CA) and Fly-Taal (FT). Unpublished manuscript, Department of Linguistics, Southern Illinois University at Carbondale.

Gilbert, G. and Makhudu, D.P. 1984b. The creole continuum in Afrikaans: A non-Eurocentric view. Unpublished manuscript, Department of Linguistics, Southern Illinois University at Carbondale.

Gullberg, M., P. Indefrey, and P. Muysken. 2009. Research techniques for the study of code-switching. In B.E. Bullock, and A.J. Toribio (eds). *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press, 21-39.

Gumperz, J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3): 663-687

Hogeweg, L. 2007. Conflicting constraints in word choice of bilinguals. Talk, Taaldag AVT Utrecht January 2007.

Hopper, P. and E.C. Traugott. 2003 [1993]. *Grammaticalization.* 2nd Edition. Cambridge: Cambridge University Press.

Janson, T. 1983. A language of Sophiatown, Alexandra and Soweto, *York papers in Linguistics* II.

Kellerman, E. 1979. Transfer and Non-Transfer: Where We Are Now. *Studies in Second Language Acquisition* 2: 37-57.

Kibogoya, A. 1995. Kiswahili/English codeswitching: Some morphological and syntactic aspects. PhD thesis, Lancaster University.

Kirsch, B. and S. Skorge with S. Magona. 2001. *Clicking with Xhosa. A Xhosa phrasebook.* Cape Town: David Philip Publishers.

Klerk, V. de. 2006. Codeswitching, Borrowing and Mixing in a Corpus of Xhosa English. *International Journal of Bilingual Education and Bilingualism* 9: 597-614

Mati, X. 2003. Using code switching as a strategy for bilingual education in the classroom. Paper presented at 21st AEAA Conference, Cape Town, South Africa.

McSwann, J. 1999. *A Minimalist Approach to Intrasentential Code Switching*. New York: Garland Press.

Muysken, P. 2000. *Bilingual speech*. Cambridge: Cambridge University Press

Muysken, P. 2011. Code-Switching. In R. Mesthrie (ed.) *Cambridge Handbook of Sociolinguistics*. Cambridge: Cambridge University Press.

Myers-Scotton, C. 1993. *Duelling languages: Grammatical structure in codeswitching.* Oxford: Oxford University Press

Myers-Scotton, C. 2006. Natural codeswitching knocks on the laboratory door. *Bilingualism: Language and Cognition* 9: 203-212

Myers-Scotton, C. and J.L. Jake. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38: 1053-1100.

Newmeyer, F.J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86: 688-695.

Poplack, S. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL. *Linguistics* 18: 581-618.

Prince, A. and P. Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell Publishers.

Schuring, G.K. 1985. *Kosmipolitiese omgangstale. Die aard, oorsprong en funksies van Pretoria-Sotho en ander koine-tale*. Pretoria: Raad vir Geesteswetenschaplike Navorsing.

Sebba, M. 2009. On the notions of congruence and convergence in code-switching. In B.E. Bullock and A.J. Toribio (eds). *The Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press, 40-57.

Senaratne, C.D. 2009. *Sinhala-English code-mixing in Sri Lanka. A sociolinguistic study.* Doctoral dissertation Radboud University Nijmegen LOT-series.

Simango, S.R. 2007. Xhosa-English codeswitching and grammatical convergence. Paper presented at ISB6, Hamburg, Germany.

Winterbach, I. 2010. *Die benedenryk*. Kaapstad/Pretoria: Human & Rousseau.

Appendix: Further English verbs with *ge-* in the Bowers (2006) corpus

(a)   en hy't sy [bəgʌ:ɜ] ge-*lug* reg oor die continent, verstaan jy (Bowers 55)
"and he lugged his baggage right across the continent, you understand"

(b)   En toe het een nou ge-*decide* dat hy gat (gaan) nou, nie een *magic mushrooms*, maar sommer ses, *but he's never done this before*. (Bowers 71)
"And so one now decided that he's going to do, not one magic mushrooms, but six, but he's never done this before."

(c)   Ja, *oh yes*, ons het mense ge-*fine* (Bowers 119)
"Yes, oh yes, we fined people"

(d)   Nou, nou… baie mense het ge-*communicate* huistoe, verstaan jy (Bowers 121)
"Now, now… many people communicated home, you understand"

(e)   en hulle't (hulle het) nou nie ge-*check* nou die map vir hoe ver goeters uitmekaar uit is 'ie (is nie) (Bowers 162)
"and they now didn't check the map as to how far apart things are"

(f)   Okay, nog 'n, nog 'n ding – ek is nie ge-*fire* nie (Bowers 390)
"Okay, another, another thing – I was not fired"

(g)   You know, the times… NAME was die een wat ons so baie ge-*fine* het (Bowers 410)
"You know, the times… NAME was the one that we fined so much"

(h)   Ja, en Suster, daar was so 'n *rumour* gewies van julle twee wat, wat 'n kamer … ge-*share* het (Bowers 421)
"Yes, and Sister, there was a rumour about you two that shared a room"

(i)   Ons het onse assets ge -ge-*split* (Bowers 852)
"We spilt our assets"

(j)   Clarence, wie ever nou iets sê na die, NAME het vir my ge-*bless*, het gesê, "voorentoe" (Bowers 858)
"Clarence, who ever says something after this, NAME blessed me, she said, 'go ahead'"