Isogloss
Open Journal of Romance Linguistics

Received: 17-12-2024
Accepted: 14-12-2025
Published: 18-12-2025

Standing on the shoulders of giants: Quantitative approaches to the identification of community norms in code-switching

Margaret Deuchar
University of Cambridge
ORCID ID: 0000-0002-3289-8981
m.deuchar@gmail.com

How to cite: Deuchar, Margaret. 2025. Standing on the shoulders of giants: Quantitative approaches to the identification of community norms in code-switching. BHL4, eds Karina Fascinetto Zago, Antje Muntendam & María Carme Parafita Couto. Special Issue of *Isogloss. Open Journal of Romance Linguistics* 11(4)/9, 1–18.
DOI: https://doi.org/10.5565/rev/isogloss.542

Abstract

In this paper I explore how collaborative work conducted with colleagues has built on previous work to arrive at an enhanced understanding not only of what code-switching is, but also how norms for its use can vary from community to community. Following an overview of our theoretical approach and of our main results I will try to trace the influence of previous work to show how key developments in twentieth century linguistics allowed us to make the progress we were able to achieve. We shall see how our current understanding has developed from pioneering early work on the notions of the speech community and linguistic repertoire, as well as on initial and subsequent attempts to formulate 'rules' of code-switching.

Keywords: code-switching, community norms, quantitative analysis, Matrix Language Frame, Muysken's typology

1. Introduction

In 1645, Isaac Newton wrote in a letter to the fellow physicist Robert Hooke: "If I've seen further, it is by standing on the shoulders of giants"[1]. This statement seems particularly relevant to the point I want to make in this paper, that insights about community norms in code-switching are only possible today because of the cumulative contributions of giants in linguistics, sociolinguistics, and psycholinguistics etc.

So how do we define code-switching? In this paper I am using the term to refer to the use of two or more languages in the same conversation, whether spoken or written (see also Deuchar 2012). Here are a couple of examples[2] taken from the Spanish/English corpus collected with colleagues in Miami, USA:

(1) bangortalk.org.uk (Miami:sastre9)

él siempre	*me*	*da*	*cumplido-s*	*así*
he always	me.DAT.1SG	give.V.PRS.3SG	compliment-PL	like this

'he always gives me compliments like this'

so I said to him, "talk to me in two more years"

(2) bangortalk.org.uk (Miami:sastre9)

pero	*hasta*	*ahora*	***because***	*él*	*nada*	*más*
but	until	now		he	nothing	more

tiene	*que*	*quince*	*añ-it-o-s*	*ya*	*para*	*dieciséis*
have.PRES.SG	than	fifteen	year-DIM-M-PL	already	for	sixteen

next month

'but until now, since he isn't more than fifteen years old, turning sixteen next month.'

Looking back on my work on code-switching with colleagues in the early part of this century, I would sum it up as gradually arriving at a realization that the code-switching patterns found in different bilingual[3] communities can be largely attributed to community-specific norms just as it is in the case of community-specific lexicons (cf. Clark, 1996: 108) or other community-specific norms (cf. Pinter 2025 on 'common knowledge'). The challenge, then, is to find appropriate ways to define these shared patterns in a way which distinguishes them from other patterns shared by other communities.

---

1 https://digitallibrary.hsp.org/index.php/Detail/objects/9792

2 Spanish words are in italics; English words in bold italics; Key to glosses: 1SG = 1st person singular; 3SG = 3rd person singular; DAT = dative; DIM = diminutive; M = masculine; NEG = negative; PL = plural; PRS = present; V = verb.

3 I follow Muysken (2013: 711) in using the term 'bilingual' to also include multilingual communities while recognising that relatively little of my research includes the latter.

2. Theoretical models

In my attempt to capture these shared patterns I have benefitted greatly from the availability of two theoretical models of code-switching developed at the beginning of this century by Myers-Scotton (2002) and Muysken (2000). Together with colleagues[4] I have been able to develop quantitative versions of these models in order to identify both uniform and variable aspects of the code-switching norms found in different bilingual communities. Our first quantitative model makes use of Myers-Scotton's Matrix Language Frame (Myers-Scotton 2002) approach and but goes beyond the analysis of individual clauses to arrive at the quantitative distribution of the matrix language in each bilingual community. The second model applies Muysken's (2000) typology of structural code-switching patterns to arrive at a quantitative result regarding the dominant type of switch in specific communities.

Myers-Scotton's (2002) clause-based model of code-switching assumes an asymmetric relation between the two (or more) languages in a clause such that one language provides the grammatical frame of the clause while the other(s) may provide lexical material belonging to a restricted range of grammatical categories. The first language is designated the matrix language (ML) of the clause, while the second is the embedded language (EL). Material in the embedded language constitutes the switches away from the matrix language. Whereas there is no restriction on the grammatical categories of constituents from the matrix language, only certain broadly 'open class' items can be drawn from the embedded language. In example (3)[5] below from our Welsh/English data, Welsh is the matrix language and English the embedded language.

(3) bangortalk.org.uk (Siarad:davies6)

oedd	'na	fath	â	ryw	***alley***	bach	yna.
be.V.3SG.IMP	there	kind	with	some		little	there

'There was kind of a little alley there.'

In identifying the matrix language in our work we draw on Myers- Scotton's System Morpheme Principle and Morpheme Order Principle (Myers-Scotton, 2002: 59) as described and exemplified by Deuchar, Webb-Davies & Donnelly (2018: 115-118). As explained in more detail by Deuchar et al. 2018, the matrix language of a clause can by identified by the language of a particular kind of 'system morpheme', e.g. an inflection on a finite verb which agrees with the subject of the verb. In addition, it can also be identified by the order of the words or morphemes in the clause. In example (3) above, the first word *oedd* is a finite verb, marked as third person singular, and it agrees with the third person subject *ryw* ***alley*** *bach* ('some little alley'). The fact that the agreement is expressed by a Welsh morpheme is an indicator (according to the System Morpheme Principle) that the matrix language of the clause is Welsh. The Morpheme Order Principle is then applied by looking at the word order of the clause. We can see that the verb comes first as in Welsh (a VSO language), and the

[4] Publications include Blokzijl et al. 2017a, Carter et al. 2011, Deuchar et al. 2007, Deuchar et al. 2016, Deuchar et al. 2018, Deuchar 2020, Deuchar accepted, et al. 2017, 2025, Herring et al. 2010, Parafita Couto et al. 2014, Parafita Couto et al. 2015, Phillips and Deuchar 2021.

[5] Welsh words are shown in italics; English words in bold italics. Glossing conventions as in fn. 2 plus IMP = imperfect.

adjective *bach* ('little') follows the noun as in Welsh (in contrast to adjective-noun order as in English). Thus according to both the System Morpheme Principle and the Morpheme Order Principle, the matrix language of the clause in example (3) is Welsh.

A similar approach was applied to clauses in our Spanish/English and Welsh/Spanish data. Example (4)[6] is from our Spanish/English data:

(4) bangortalk.org.uk (Miami:herring9)
my mom got ***the*** *manguera*
hosepipe
'My mum got the hosepipe.'

Since Spanish and English are both SVO languages, the Morpheme Order Principle is not so useful in identifying the matrix language of clauses with code-switching. The System Morpheme Principle, however, is more useful. If we look for the finite verb with subject-verb agreement, we find the English finite verb *got*. Therefore, we can tell that (4) has English as a matrix language.

Finally, we provide an example (5)[7] below from our Spanish/Welsh data (see bangortalk.ac.uk). As Spanish is an SVO language while Welsh is VSO, we can make use of the Morpheme Order Principle to identify the matrix language in (5).

(5) bangortalk.org.uk (Patagonia:patagonia7)

fasen	nhw	isio	nhw
be.V.3P.PLUPRF.SM	they.PRON.3P	want.N.MSG	they.PRON.3PL

i-(y)r	*campamentos*
to.PREP-the.DET.DEF	camp.N.M.PL

'They'd need them for the camps.'

In example (5) above the finite verb *fasen* appears before the subject *nhw* in line with VSO order, thus suggesting that Welsh is the matrix language. Furthermore, applying the System Morpheme Principle we can identify the third person subject-verb agreement on the verb *fasen* and the subject *nhw* 'they'. The open class Spanish word *campamentos* is inserted in the Welsh morphosyntactic frame. Thus Welsh is the matrix language of example (5) and Spanish is the embedded language.

3. Quantitative analysis of the distribution of the matrix language.

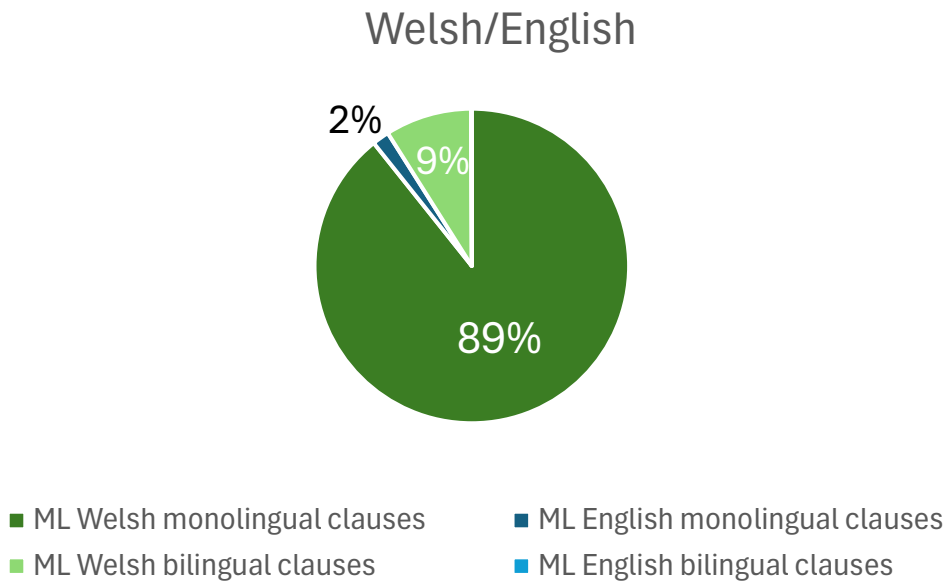
As outlined above, not only did we apply Myers-Scotton's MLF to individual clauses following her methods, but we developed her approach from a qualitative to a quantitative method, applying her model to all finite clauses in each set of data analysed. This involved extracting all the finite clauses in each dataset and coding them according to (1) monolingual or bilingual status (i.e. containing items from a single language or containing code-switching) and (2) matrix language of clause. This

6 Glossing conventions as in fn. 2, plus DEF = definite, DET = determiner.

7 Glossing conventions as in fn. 2, plus PLUPERF = pluperfect; PRON = pronoun; SM = soft mutation.

allowed us to arrive at the quantitative distribution of (1) monolingual vs. bilingual clauses and (2) matrix language in the data from each community. On the basis of the analysis of a large number of clauses (e.g. 66,429 in our Welsh/English data) we went on to argue that our quantitative results could be said to represent the communicative norms of the communities. Our detailed results are available in the publications listed in footnote 4 but Figures 1-3 below summarise them and represent pictorially our perception of the norms in the extent of code-switching as well as the predominant choice of the matrix language in each community.

Figure 1. Matrix language distribution in the Welsh/English data



Source: Deuchar *et al.* (2018: 89)

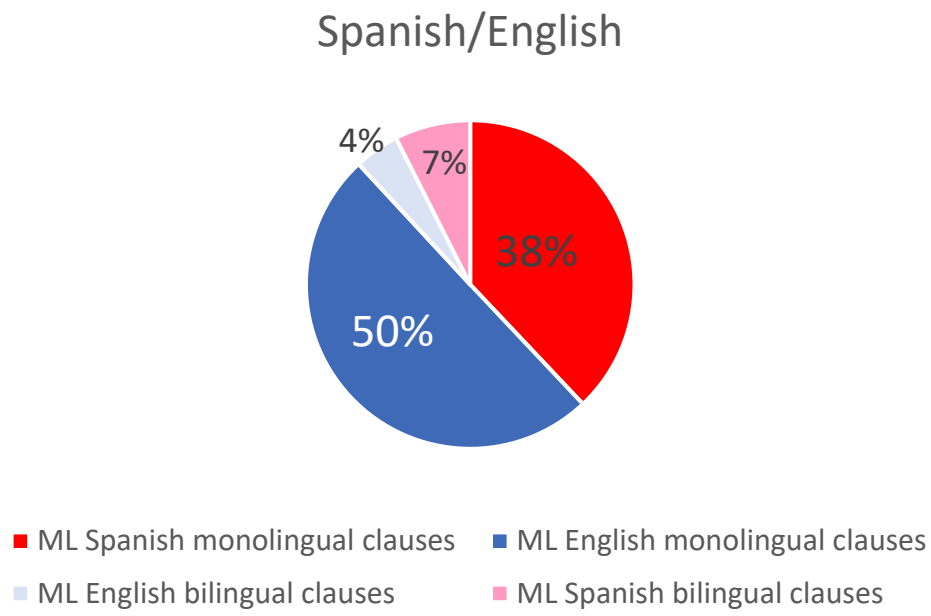
As a result of identifying all finite clauses as either monolingual or bilingual, we were then able to calculate the overall proportion in each dataset of bilingual clauses or in other words, the proportion of clauses that contained code-switching. This gave us an idea of how frequently bilingual speakers in each community code-switch within the clause. Once this had been calculated for each community we could make cross-community comparisons regarding the normal frequency of code-switching.

Starting with the data represented in Figure 1, we can see that 9% of the clauses in the Welsh/English data 9% are classified as bilingual and so include code-switching, whereas the rest (91%) are monolingual. We can interpret these numbers as indicating that Welsh/English bilinguals may code-switch about 9% of the time on average when they are speaking to other bilinguals.

If we compare the results depicted in Figure 1 with the percentage of bilingual clauses found in the Spanish/English data (see Figure 2), we see that it is higher, in fact just under 12% (7.46% + 4.43%). Finally, Figure 3 shows that the percentage of bilingual clauses found in the Welsh/Spanish data is lower than in both of the other corpora, in fact 5.47% (4.94% + 0.53%). Taken together, these results can be interpreted as showing that the norms regarding the frequency of code-switching can

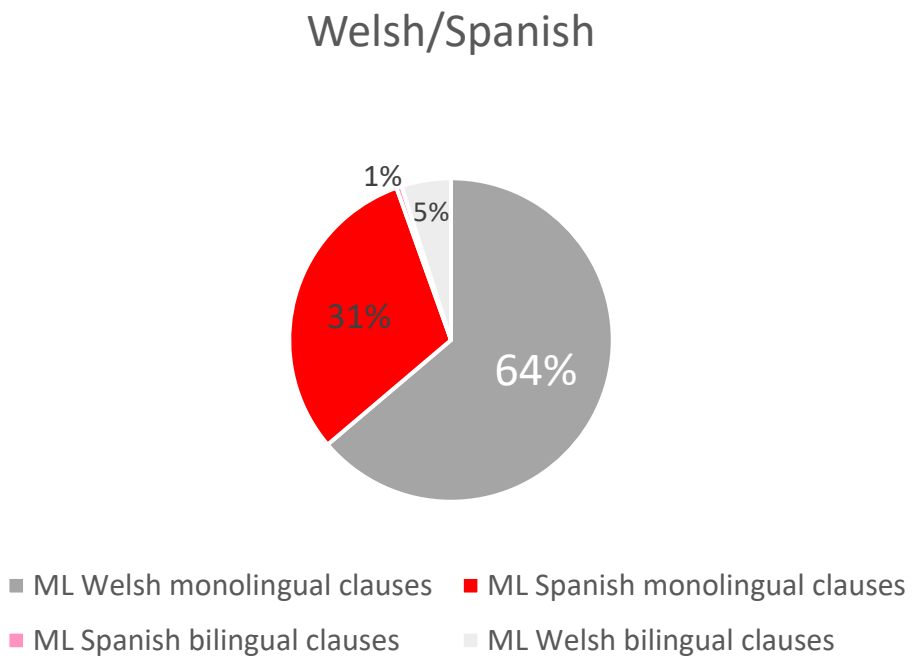
vary from community to community and that these norms are arguably internalized by bilingual speakers on the basis of the linguistic input they have received while growing up and living in their community.

Figure 2. Matrix language distribution in Spanish/English data



Source: Blokzijl et al. (2017a, b), Deuchar (2020)

Figure 3. Matrix language distribution in Welsh/Spanish data



Source: Deuchar (accepted), Somsen (2021)

In addition to coding the finite clauses in our data as either monolingual or bilingual, we also identified the matrix language of all clauses in order to provide an overall impression of the extent to which the choice of the matrix language is uniform for a community or variable. Looking first at the results for the Welsh/English data we can see that there is a high degree of uniformity in the choice of Welsh in both monolingual and bilingual clauses. 98% of the monolingual clauses in the data are entirely in Welsh, and 99% of the bilingual clauses have Welsh as their matrix language. This reflects the normal practice exhibited by Welsh/English bilinguals to use Welsh as the grammatical frame in most of their utterances, even when they code-switch and insert English words and phrases.

Figure 2, when compared with Figure 1, shows quite different norms regarding the choice of the matrix language in the Spanish/English bilingual community of Miami[8]. Whereas in the Welsh/English data shown in Figure 1 we saw that Welsh is overwhelmingly the matrix language in both bilingual and monolingual clauses, in the Miami data we see that English is the preferred (matrix) language for monolingual clauses, i.e. when no code-switching is taking place, whereas Spanish is preferred as the matrix language where code-switching is involved, as in bilingual clauses. Of the total number of 6232 clauses in the sample of the Miami data analysed, 2366, or exactly 50% are monolingual English clauses. But despite the preference for English in monolingual clauses, English appears as the matrix language of less than half of the 741 bilingual clauses, in which Spanish is the matrix language 63% of the time. This is an interesting result which suggests that although bilinguals in Miami may decide to speak English more often than Spanish to each other, when they do choose Spanish they are quite likely to switch to English in the same clause. The frequent use of English monolingual utterances between bilinguals in Miami differs from the norm among bilinguals in Wales, where monolingual English tends to be reserved for communication with English monolinguals. So in both bilingual and monolingual clauses there is less uniformity in the choice of the matrix language in Miami compared with the Welsh/English community. In the Welsh/English community the same matrix language (Welsh) is highly preferred in both bilingual and monolingual clauses, whereas in Miami, English is preferred in monolingual clauses and Spanish (relatively) in bilingual clauses.

Looking at the results in Figure 3 on the distribution of the matrix language shown in the Welsh/Spanish data in Figure 3, we see that Welsh is the predominant matrix language in both monolingual and bilingual clauses. This parallels the situation shown in Figure 1 for Welsh/English, with the difference that the matrix language is not quite so uniform in Patagonia. Whereas we have seen that there is relatively little use of monolingual English between bilinguals in Wales, in Patagonia the corresponding majority language, Spanish, is used in 32% of monolingual clauses in communication between bilinguals. We may also note that the matrix language of the bilingual clauses in the Patagonia data is less uniform than in the Welsh/English data, with Welsh being the matrix language in 90% of the bilingual clauses and Spanish appearing as matrix language in 10%.

[8] Note that the Miami data used for the analyses differed in quantity from the datasets from the other two corpora. Whereas the entire corpora of the Welsh/English and Welsh/Spanish data were analysed, the clauses extracted from the Miami corpus were limited to those containing DPs (phrases consisting of nouns and determiners) which had been extracted for another project (see Blokzijl et al 2017a).

We have suggested that our results regarding the frequency of code-switching and the distribution of the matrix language in the speech of bilinguals in different bilingual communities reflect the norms of code-switching in which members of those communities are socialised as they develop. If, as we suggest, code-switching norms are actually input to children in the process of bilingual acquisition, we may expect these norms to emerge in their early speech. Little research has been done on this topic to date, but Phillips & Deuchar (2021) conducted a study of the data in the CHILDES CIG1 corpus, a longitudinal study of Welsh/English developing bilinguals (see Aldridge et al. 1997). We predicted that the code-switching patterns previously identified in the Welsh/English bilingual community discussed above would already be discernible in this corpus of child data from seven child speakers aged between 1;9 and 2;6. Our analysis focused specifically on child utterances consisting of two words in which one word was Welsh and the other English. Using a method of identifying the matrix language which we adapted for child data, we analysed 172 child clauses. We found that just as in the adult data, the vast majority of the utterances had a Welsh matrix language or morphosyntactic frame. Furthermore, a detailed analysis of the specific input to one of the children in the corpus showed a very similar pattern. We concluded that the code-switching patterns in the linguistic input to children begin to be reproduced in child productions from a very young age.

4. Quantitative analysis of code-switching patterns

So far we have suggested that code-switching practices particular to bilingual communities have their own norms which are known and applied by community members. These norms will include the amount of code-switching that is conventional in the community as well as norms in the choice of matrix language or grammatical frame for each clause. In addition, Muysken (2000: 249) has argued that community norms can be identified in "quantitative tendencies towards particular patterns" of code-switching. He identifies three main alternative patterns which he labels 'insertion', 'alternation' and 'congruent lexicalisation'.

Insertion involves a switch from one language into that which provides the main grammatical frame and is exemplified in a Swahili/English example from Myers-Scotton (1993:86) shown in (6) below:

(6) Deuchar et al. (2007: 303)

a-na-ku-l-a	***plate***	m-bili	z-a	murram
3SG-PRS-NFIN--eat-IND		CLM10-two	CLM10-of	maize

'He eats two plates of maize.'

As they explain, an English word *plate* is inserted into a Swahili grammatical frame with Swahili word order and Swahili morphology. In the alternation pattern, as Muysken (2000:7) describes it, "a constituent from language A…is followed by a constituent from language B". Deuchar et al. (2007: 304) select a clear example of alternation from Poplack (1980: 594) as shown in example (7) below, in which English is shown in bold:

(7) *si tú eres* *puertorriqueño*, ***your father's a Puerto Rican***,
if 2SG be.PRS.2SG Puerto-Rican

you should at least *de vez en cuando,* ***you know***,
from time to time

hablar *español*
speak.NFIN Spanish
'If you're Puerto-Rican, you're father's a Puerto Rican, you should at least sometimes speak Spanish.'

As for the third pattern of code-switching, congruent lexicalisation, Muysken describes this in a more recent publication (Muysken 2013: 713) as "the use of elements from either language in a structure that is wholly or partly shared by languages A and B". He provides an example from Poplack (1980: 589) that I have shown below in (8) as in Poplack's paper, with Spanish in capitals followed by English translation in parentheses. I have added clause boundaries.

(8) [[Why make Carol SENTARSE ATRAS] [PA' QUE (sit in the back so) everybody has to move] [PA' QUE SE SALGA]] (for her to get out)?

In this example we see switches from one language to another at points where Spanish and English have shared structures, with the first stretch of Spanish even crossing a clause boundary (cf. Muysken 2013: 713).

With reference to his features, Muysken makes some suggestions regarding the predominant pattern in some published data on code-switching. For example, he compares data from Pfaff (1979) and Poplack (1980) to suggest that insertion is more predominant in Pfaff's data than in Poplack's. However, he apparently did not have access to the full corpora, and suggests that a more exhaustive treatment would be preferable.

The tools for a more exhaustive treatment are in fact provided in a detailed list of diagnostic features for each code-switching pattern provided by Muysken (2000:230). These can be tested one by one against switches in data from a particular community. Deuchar, Muysken and Sung-lan Wang (2007) extracted the switches from three corpora to which they had access: Welsh/English, Tsou/Mandarin Chinese and Taiwanese Chinese/Mandarin Chinese. Intraclausal switches were defined as material in the embedded (non-matrix) language. Deuchar et al. (2007) developed a system of scores for each of Muysken's three patterns such that a score was given to each switch in three bilingual corpora to which they had access. Where the value of a feature matched that expected for particular pattern, a score of '1' was given for that pattern. Non-matching scores were given a score of '-1', however, and the scores for each pattern were then totalled. The highest positive score was taken to indicate the predominant pattern or norm in that community. Further details of the application of the scoring system, including examples, are described by Deuchar et al. (2007:308-335). Their results showed that there are indeed community-specific norms in terms of a predominant code-switching pattern. For Welsh/English this was insertion, for Taiwanese/Mandarian it was congruent lexicalisation, and for Mandarin/Tsou it was

both insertion and congruent lexicalisation. For Welsh/English there was also a secondary pattern, of congruent lexicalisation.

So far we have seen considerable evidence for community-specific norms in code-switching. Through quantitative development of Myers-Scotton's MLF and Muysken's diagnostic features we have seen that community norms can be identified in terms of at least (a) quantity of code-switching; (b) distribution of the matrix language; and (c) predominant code-switching patterns. While our specific work has built on the pioneering work of Myers-Scotton and Muysken, I now want to point out how its foundations can be traced back to the insights of some twentieth century linguists, starting in the 1950s.

5. Foundational insights of twentieth century linguists

5.1. 1950s

Uriel Weinreich, working in the 1950s, is probably responsible for the establishment of the area of language contact as a focus of academic study His main interest was 'interference', which he defined as involving "those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich, 1953, 1968: 1). In his book *Languages in Contact*, he said that "the ideal bilingual switches from one language to the other according to appropriate changes in the speech situation (interlocutors, topics, etc.), but not in an unchanged speech situation, and certainly not within a single sentence" (Weinreich, 1953, 1968: 73). In this statement we can see that he recognised what was later called 'extrasentential switching' by Poplack (1981: 599) but not intrasentential or intraclausal switching (cf. Deuchar 2012). Nevertheless, he states elsewhere that "There is some reason to believe that a facility in switching languages even within a single sentence or phrase is characteristic of some bilinguals" (Weinreich 1953, 1968: 68). Weinreich provides an example of an utterance in Yiddish in which a relative clause contains the English verb *accumulate* but with an inflection as in Yiddish. But in general he assumed that languages should be kept structurally separate and in relation to this example of switching he speculated as to whether the speaker might be involved in a process of language shift. He did not consider the possibility that code-switching might be a community norm, but saw it as an individual type of linguistic behaviour.

5.2. 1960s

Another giant academic who had a strong influence on the study of code-switching from the 1960s onwards was John Gumperz. He introduced into the fledgling field of sociolinguistics the notion of speech community, which most code-switching researchers now assume as a given. Gumperz defines 'speech community' as "any human aggregate characterized by regular and frequent interaction over a significant span of time and set off from other such aggregates by differences in the frequency of interaction" (Gumperz, 1964: 137). Gumperz says that studying the language(s) of the speech community allows us to identify that community's verbal repertoire, or "the totality of linguistic forms regularly employed in the course of socially significant

interaction" (Gumperz, 1964: 137). Nowadays this is normally referred to as the community's 'linguistic repertoire'. The notions of speech community and linguistic repertoire were new in the 1960s but widely used by scholars nowadays and provided the basis for the recognition by linguists that community-specific norms may exist.

Gumperz (1964) was also the author of one of the very first studies of code-switching in a specific community, Hemnesberget in northern Norway. In his analysis of the data he collected, he differentiated between the two types of code switching that he was able to observe. Switching took place between two varieties of Norwegian, Bokmål one of the two standard languages and a dialect, Ranamål, which is spoken in northern Norway. He identified two types of code-switching that he called 'transactional' and 'personal'. Transactional switching, which he later called situational switching, referred to a change in language according to the situation. For example, Bokmål was used in formal education and church services in Hemnesberget, but Ranamål was used in informal conversation. Gumperz also found that switching within the same situation was practised by a group of students who studied elsewhere but who were currently at home in Hemnesberget on vacation. These students used Ranamål and Bokmål in the same sentence during informal discussion, in a phenomenon he initially called 'personal' or 'metaphorical' and, later, 'conversational code-switching'. Consider example (9).

(9) Blom & Gumperz (1972: 429)[9]

de	*voel*	du	***mellom***	***en***	***faemm***	***saeks***
that	choose	you	among	the	five	six

'You choose that from among five or six.'

Gumperz pointed out that linguistic varieties such as Bokmål and Ranamål are defined by their different co-occurrence restrictions. He said that these restrictions allow us to "segment verbal repertoires into distinct varieties" (Gumperz 1964: 140). However, he recognised that code-switching of the kind illustrated by example (2) involved a loosening of co-occurrence restrictions, resulting in an "erosion of the linguistic boundary between Ranamål and Bokmål" (Blom and Gumperz 1972: 429). As we shall see, later work on intraclausal code-switching according to the Matrix Language Frame approach (Myers-Scotton 2002) would demonstrate that even the relaxing of co-occurrence restrictions can be subject to constraints or rules.

Nevertheless, what Gumperz was really interested in was not what kinds of words could be switched or what the switch point might be, but rather in the social meaning of code-switching. He argued that the students were using both Ranamål and Bokmål to represent their shared identity: both with the local community and with their identity with higher education.

While Gumperz's main interest in code-switching was on its function, he also recognised that code-switching could occur both between and within sentences (Gumperz 1977: 1-2), an important insight that was to be built upon by later work on the structure of code-switching.

Dell Hymes was a colleague of Gumperz whose work also contributed to laying the foundation of later work in code-switching. In 1966 Hymes introduced the concept of 'communicative competence' at a conference and later published several articles

[9] *Italics* indicate words in Ranamål and ***bold italics***, words in Bokmål.

about it (see e.g. Hymes 1972). This notion involved an elaboration and a critique of Chomsky's notion of competence, by which Chomsky meant knowledge of the rules of grammar (cf. Chomsky 1965: 4). Hymes accepted that speakers need to have knowledge of grammar in order to communicate, but in his view speakers also needed to have knowledge of how to use those rules. This knowledge would include knowledge of what is appropriate in the speaker's community, and what the norms of use are. Hymes thus recognised community-specific norms, laying the foundation of a focus on these norms as part of code-switching research.

5.3. 1970s

Despite the pioneering work of Gumperz and Hymes in the 1960s, relatively little was understood about the structural nature of code-switching until the end of the 1970s. It was then that Pfaff (1979) published an influential article called 'Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English' in *Language*. She used a corpus with Spanish English data from Mexican Americans in Southwest USA, and drew various conclusions regarding when switching can occur and when it cannot. For example, Pfaff found that switches are favoured when the surface structures are similar in both languages, as in an auxiliary followed by a verb, shown in example (10) below.

(10) *Estaba* training *para pelear* ('he was training to fight'). (Pfaff 1979: 299)

In NPs Pfaff found that switches between a determiner and a noun as in (11) are the most frequent types.

(11) *El* flight *que sale de Chicago* ('the flight which leaves from Chicago') (Pfaff 1979: 305)

As Pfaff pointed out, there is no structural conflict here, since the order of the determiner and the noun are the same in both English and Spanish. Pfaff suggests that structural conflict does however arise for adjectives juxtaposed with nouns in NPs, since adjectives generally precede the noun in English but follow it in Spanish. For this reason she suggests that adjective switches are "limited within the NP"[10] but "are unrestricted when they take the form of predicate adjectives" (Pfaff 1979: 314). So as she shows in relation to example (12), the English adjective can occur in the predicate as there is no structural conflict between English and Spanish there.

(12) No están *free* ('they're not free'). (Pfaff 1979: 305)

Pfaff was one of the first scholars to identify an innovation in code-switching which cannot be attributed to the grammar of one of the languages alone, but which involves the use of a finite form of the Spanish verb *hacer* ('do' or 'make') juxtaposed with an English non-finite verb form. Pfaff's example is shown below as no. (13) and she describes it as rare.[11]

[10] This insight on Pfaff's part has received recent empirical support in a paper by Balam and Parafita Couto (2019).

(13) *Su hija hace* teach *allá en San José* ('His daughter teaches there in San Jose'). (Pfaff 1979: 301)

This innovation has received more attention recently (cf. Vergara Wilson & Dumont 2013, Balam et al. 2020).

5.4 1980s

Pfaff's work on code-switching constraints was followed by Poplack's (1980) seminal paper entitled 'Sometimes I'll start a sentence in Spanish y termino en español'. This paper may be considered one of the most influential articles on code switching of the 20th Century. In this paper she pursues a similar aim to Pfaff in that she proposes various constraints on code-switching. One of the best known is the 'equivalence constraint'. According to this, a switch is possible only if it involves no violation of the surface syntactic rules of either language. This is illustrated in the example of possible switch points shown in (14). (See Poplack 1980: 586).

(14) ENG: I | told him | that | so that | he | would bring it | fast
SPA: *(Yo)* | *le dije* | *eso* | *pa' que* | *(él)* | *la trajera* | *ligero*
CS: I told him that *pa' que la trajera ligero*

In (14), the vertical lines show when you can make a switch, whereas in between the vertical lines switching is not possible. For example, a switch from English to Spanish would be possible after 'told him', but not between 'told' and 'him'. This is because the placement of the object pronoun is after the verb in English, but before it in Spanish. For that reason, switching between the two words is not possible.

Since the publication of Poplack's (1980) paper, various counter examples have been proposed (cf. Berk-Seligson, 1986, Sciullo, Muysken and Singh, 1986, Muysken, 1995: 192-196, Deuchar, 2020: 5-6. Nowadays, we think more of what she called the 'equivalence constraint' as a kind of a principle or a facilitation factor (cf. Poplack 2001).

In the same decade as Poplack's work on the structural nature of code-switching, Joshi (1985) introduced the notion of the matrix language in a study of Marathi/English code-switching. In his data he found an asymmetry between the two languages such that Marathi, the matrix language, provided the grammatical frame into which open class items from English, the embedded language, could be inserted. He formulated the relation between the two languages in terms of an asymmetrical "switching rule" which restricted the "switchability" of closed class items.

5.5. 1990s

When Myers-Scotton (1993) introduced the first version of her Matrix Language Frame model, it clearly owed a great deal to Joshi's insights on the asymmetric relation between the matrix and embedded language. Furthermore, her distinction between system and content morphemes mirrors Joshi's distinction between closed and open class items, which she retains in her later and more nuanced "4-M model" (Myers-Scotton 2002:16-19).

6. Conclusion

In this paper I have tried to show how my work with colleagues on the quantitative identification of code-switching norms in several bilingual communities was a development of the seminal work of Myers-Scotton (2002) and Muysken (2000) and also built on work by prominent linguists in the second half of the twentieth century. I have argued that Weinreich was one of the first to identify the phenomenon of code-switching and that Gumperz not only introduced the notions of speech community and linguistic repertoire but also applied them to one of the first descriptive studies of code-switching in a community. Hymes was then one of the first to criticize Chomky's notion of competence as excluding communicative norms, thus opening the way for these norms to be studied. Finally, Pfaff, Poplack and Joshi among others initiated the structural study of code-switching, laying foundations for the work of Myers-Scotton and Muysken. These approaches in turn provided the basis for our quantitative models, and other scholars have gone on to use our models on other data (cf. e.g. Forker 2019, Hofweber et al. 2020, Lipski 2016, Parafita Couto & Stadthagen 2019, Treffers-Daller et al. 2022).

Acknowledgments

I would like to thank all my collaborators, Fraibet Aveledo, Jeffrey Blokzijl, Diana Carter, Kevin Donnelly, Jon Herring, Siân Lloyd-Williams, Pieter Muysken, M. Carmen Parafita Couto, Sharon Phillips, Caroline Piercy, Elen Robert, Jonathan Stammers, Jorge Valdés Kroff, Sung-lan Wang and Peredur Webb-Davies. I would also like to acknowledge the valuable contributions by Bangor University, the Arts and Humanities Research Council, and the Economic and Social Research Council in the United Kingdom. Thanks are due to Aaron Santa María for editorial assistance with this paper.

References

Aldridge, Michelle, Borsley, Robert, Clack, Susan, Creunant, Gwen & Jones, Bob M. 1997. The acquisition of noun phrases in Welsh. In Sorace, A., Heycock, C., & Shillcock, R. (eds.) *Proceedings of the GALA '97 Conference on Language Acquisition, Edinburgh 4-6 April 1997*, 6-9. Edinburgh: Human Communication Research Centre.

Balam, Osmar, Parafita Couto, María del Carmen & Stadthagen-González, Hans (2020). Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism*, *24*(5-6), 952-967.https://doi-org.ezp.lib.cam.ac.uk/10.1177/1367006920911449

Balam, Osmar, & Parafita Couto, María del Carmen. 2019. Adjectives in Spanish/English code-switching: Avoidance of grammatical gender in bi/multilingual speech. *Spanish in Context* 16(2): 194-216.

Berk-Seligson, Susan. Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15(3) 313-348.

Blokzijl, Jeffrey, Deuchar, Margaret, & Parafita Couto, María del Carmen. 2017a. Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages* 2(4): 20. https://doi.org/10.3390/languages2040020

Blokzijl, Jeffrey, Deuchar, Margaret, Donnelly, Kevin, Mauder, El, and Parafita Couto, María del Carmen. 2017b. How can determiner asymmetry in mixed nominal constructions inform linguistic theory? Paper presented at the Meeting of Societas Linguistica Europaea, Zurich, Switzerland, September 10–13

Blom, Jan-Petter & John J.Gumperz. 1972. Social meaning in linguistic structures: code-switching in Norway. In J.J. Gumperz & Dell Hymes (eds.) *Directions in Sociolinguistics: The Ethnography of Communication*, 407-434. New York: Holt, Rinehart and Winston.

Carter, Diana, Davies, Peredur, Deuchar, Margaret, & Parafita Couto, María del Carmen Parafita Couto. 2011. A systematic comparison of factors affecting the choice of matrix language in three bilingual communities. *Journal of Language Contact*, *4*(2), 153-183.

Clark, Herbert H. 1996. Using Language. Cambridge: Cambridge University Press.
Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: The MIT Press.

Deuchar, Margaret. 2012. Code-switching. In Chapelle, C. (ed.) *Encyclopedia of Applied Linguistics*, 657-664. New York: Wiley.

Deuchar, Margaret, Muysken, Pieter, & Wang, Sung-Lan 2007. Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 298-340.

Deuchar, Margaret, Webb-Davies, Peredur, & Donnelly, Kevin. 2018. Building and using the Siarad Corpus: Bilingual conversations in Welsh and English (Vol. 81). John Benjamins Publishing Company.

Deuchar, Margaret, Donnelly, Kevin & Piercy, Caroline 2016. 'Mae pobl monolingual yn minority': Factors favouring the production of code-switching by Welsh/English speakers. In Durham, M. & Morris, J. (eds.) *Sociolinguistics in Wales*, 209-39. London: Palgrave Macmillan.

Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5: 22.

Deuchar, Margaret. accepted. Language contact and corpus linguistics. In Cerno, L. Döhla, H.J., Gutiérrez Maté, M., Hesselbach, R. & Steffen, J. (eds.), Contact varieties of Spanish and Spanish-lexified contact varieties. Berlin & New York: Mouton de Gruyter.

Eppler, Eva D., Luescher, Adrian, & Deuchar, Margaret. 2017. Evaluating the predictions of three syntactic frameworks for mixed determiner–noun constructions. *Corpus Linguistics and Linguistic Theory*, 13(1), 27-63.

Eppler, Eva D., Luescher, Adrian, & Deuchar, Margaret. 2025. Testing the Predictions of Word Grammar, the Minimalist Programme and the Matrix Language Frame Model for German/English Mixed Determiner–Noun Constructions. In Duran Eppler, Eva, Gisborne, Nikolas & Rosta, Andrew (eds.), *Word Grammar, Cognition and Dependency*, 252-275. Cambridge: Cambridge University Press.

Forker, Diana. 2019. Sanzhi-Russian code switching and the Matrix Language Frame model. *International Journal of Bilingualism* 23(6): 1448–1468 https://doi.org/10.1177/1367006918798971

Gumperz, John. 1964. Linguistic and social interaction in two communities. *American Anthropologist* 66(6): 137-153. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00100

Gumperz, John. 1977. The sociolinguistic significance of conversational code-switching. *RELC Journal* 8(2): 1-34. https://doi.org/10.1177/003368827700800201

Herring, John R., Deuchar, Margaret, Parafita Couto, María del Carmen & Moro Quintanilla, Monica. 2010. Evaluating predictions about codeswitched determiner-noun sequences using Spanish/English and Welsh/English data. *International Journal of Bilingual Education and Bilingualism* 13: 553–73.

Hofweber, Julia, Marinis, Theodoros & Treffers-Daller, Jeanine. 2020. How different code-switching types modulate bilinguals' executive functions - a dual control mode perspective. *Bilingualism: Language and Cognition* 23 (4),909-925. https://doi.org10.1017/S1366728919000804

Hymes, Dell. 1972. On communicative competence. In Giglioli, P.P. (ed.) *Language and Social Context: Selected Readings*, 269-293.

Joshi, Aravind K. 1985. Processing sentences with intrasentential code switching. In Dowty, D.R., Kartunnen, L. & A. Zwicky (eds.) *Natural Language Parsing,* 190-205. https://doi.org/10.1017/CBO9780511597855.006

Lipski, John M.. The role of unintentional/involuntary codeswitching: Did I really say that? In Guzzardo Tamargo, R. E., Mazak, C.M. & Parafita Couto, María de la Carmen (eds) Spanish-English Codeswitching in the Caribbean and the US, 139-168. https://doi.org/10.1075/ihll.11.06lip

Muysken, Pieter. 1995. Code-switching and grammatical theory. In Milroy, L. & Muysken, P. (eds.) *One Speaker, Two Languages*, 177-198. Cambridge: Cambridge University Press.

Muysken, Pieter. 2000. Bilingual speech: A typology of code-mixing. Cambridge: Cambridge University Press.

Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16 (4), 709-390. https://doi.org/10.1017/S1366728912000727

Myers-Scotton, Carol. 1993. Common and uncommon ground: Social and structural factors in codeswitching. *Language in Society* 22(4): 475-503. https://doi.org/10.1017/S0047404500017449

Myers-Scotton, Carol. 2002. Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford University Press.

Parafita Couto, María del Carmen, Davies, Peredur, Carter, Diana & Deuchar, Margaret. 2014. Factors influencing code-switching. In Thomas, E. & Mennen, I. (eds). *Advances in the Study of Bilingualism*, 111-138. Clevedon: Multilingual Matters.

Parafita Couto, María del Carmen, Deuchar, Margaret & Fusser, Marika. 2015. How do Welsh/English bilinguals deal with conflict? Adjective-noun resolution. In Stell, G. & Yakpo, K. (eds.) *Code-Switching between Structural and Sociolinguistic Perspectives,* 65-84.

Parafita Couto, María del Carmen & Stadthagen-González, Hans. 2019. *El book* or *the libro*? Insights from acceptability judgments into determiner/noun code-switches. *International Journal of Bilingualism* 23 (1): 349-360.

Pfaff, Carol. 1979. Constraints on Language Mixing: Intrasentential Code-Switching and Borrowing in Spanish/English. *Language* 55(2): 291–318. https://doi.org/10.2307/412586

Phillips, Shannon, & Deuchar, Margaret. 2021. The Role of the Input in the Acquisition of Code-Switching. In Røyneland, U. & Blackwood, R. (Eds.) *Multilingualism across the Lifespan*, 56-80. Routledge. https://doi.org/10.4324/9781003125815-3

Pinker, Steven. 2025. When everyone knows that everyone Knows…: common knowledge and the science of harmony, hypocrisy and outrage. London: Allen Lane Penguin Books.

Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics* 18 (7-8): 581-618. https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, Shana. 2001. Code-switching (linguistic). *International Encyclopedia of the Social and Behavioral Sciences* 12: 2062-2065.

Sciullo, Anne-Marie Di, [Muysken, Pieter, & Singh, Rajendra. 1986. Government and code-mixing. *Journal of Linguistics*, *22*: 1-24.

Somsen, Anna. 2021. Codeswitching patterns in the speech of Welsh-Spanish bilinguals. Unpublished Master's thesis, Utrecht University.

Treffers-Daller, Jeanine, Majid, Sheikha, Yap, Ngee T. & Flynn, Naomi. 2022. Explaining the Diversity in Malay-English Code-Switching Patterns: The Contribution of Typological Similarity and Bilingual Optimization Strategies. *Language* 7, 299. https://doi.org/10.3390/languages7040299

Weinreich, Uriel. 1953. Languages in contact: Findings and problems. New York: Linguistic Circle of New York.

Weinreich, Uriel. 1968. Languages in contact: Findings and problems. The Hague; Paris: Mouton & Co.

Vergara Wilson, Damián & Dumont, Jenny. 2013. The emergent grammar of bilinguals: The Spanish verb *hacer* 'do' with a bare English infinitive. *International Journal of Bilingualism* 19(4) 444-458. https://doi.org/10.1177/1367006913516047