BRILL

Journal of Language Contact 4 (2011) 153–183

JOURNAL OF LANGUAGE CONTACT

brill.nl/jlc

A Systematic Comparison of Factors Affecting the Choice of Matrix Language in Three Bilingual Communities

Diana Carter, Margaret Deuchar, Peredur Davies, María del Carmen Parafita Couto

ESRC Centre for Research on Bilingualism in Theory and Practice
Bangor University
37-41 College Road, Bangor, Gwynedd, Wales, UK. LL57 2DG
diana@dianacarter.ca

Abstract

In this paper we compare the code-switching (CS) patterns in three bilingual corpora collected in Wales, Miami and Patagonia, Argentina. Using the Matrix Language Framework to do a clause-based analysis of a sample of data, we consider the impact of structural relationships and extra-linguistic factors on CS patterns. We find that the Matrix Language (ML) is uniform where the language pairs have contrasting word orders, as in Welsh-English (VSO-SVO) and Welsh-Spanish (VSO-SVO) but diverse where the word order is similar as in Spanish-English (SVO-SVO). We find that the diversity of the ML in Miami is related to the diversity of degrees of proficiency, ethnic identities, and social networks amongst members of that community, while the uniformity of the ML in Wales is related to the uniformity of these factors. This is not so clear in Patagonia, however, where there is little CS produced in conversation. We suggest that the members of the speech community use Spanish or Welsh mostly in a monolingual mode, depending on the interlocutor and the social situation.

Keywords

matrix language, code-switching, Spanish, English, Welsh

1. Introduction

Research on code-switching (CS) since the 1980s has been an area of intensive study and debate. The general aim of linguists working in this area has been to account for the regularities found in CS patterns. Poplack (1980) was one of the first to collect large amounts of data from a single community, and to propose two potential universal constraints on CS, the "equivalence constraint" and the "free morpheme constraint" (cf. section 2). Although these appeared to account well for her own Spanish-English data, many counter-examples

DOI 10.1163/187740911X592808

were found in other language pairs (see e.g. MacSwan, 2009) and to this day linguists are trying to account in a systematic way for the great range of diversity in CS patterns while pursuing the idea that some regularities can be found.

In order to achieve this goal we argue that an analysis of more than one CS corpus is necessary, preferably with the same language found in more than one contact setting, and ideally with the languages involved belonging to more than one structural type. Furthermore, given that extra-linguistic factors play a role in CS patterns, researchers need access to information about those characteristics of speakers which seem to be shared as community norms, as well as those which vary from speaker to speaker. A comparison of CS in different language combinations between different communities will allow us to flesh out the contribution of both typological and extra-linguistic factors on bilingual speech patterns.

The study on which we will report meets the above criteria, in that we have collected three bilingual corpora from Wales, Miami and Patagonia involving three languages (Welsh, Spanish, English) of two structural types (VSO and SVO) in three different combinations (Welsh-English [VSO-SVO], Spanish-English [SVO-SVO], Spanish-Welsh [SVO-VSO]). In addition, all participants have completed questionnaires which provide information about a range of factors including age, gender, proficiency, identity, and social network.

In this paper we shall address the question of how clause-based CS patterns may be related to both structural and extra-linguistic factors in the three communities. Our analysis of CS patterns will make use of Myers-Scotton's (2002) Matrix Language Frame (MLF) model for reasons outlined in the next section and will draw on the results from our questionnaires in assessing the impact on CS patterns of proficiency in the two languages, identity and social network.

2. Selecting a model for the analysis of CS patterns

In order to analyze the code-switching patterns in the three corpora we considered that the model needed to be able to (1) deal with production data; (2) provide a clause-based analysis; and (3) cover both monolingual and bilingual clauses. First, a model that can deal with production data is clearly necessary for a corpus-based study, and because of the negative attitudes to CS in some communities we know that an approach based entirely on judgments as established in generative models might be influenced by prescriptive attitudes. Second, we favor a clause-based analysis because it is switching within the clause that has proved to be the most interesting to linguists due to the

challenges it poses for any syntactic theory. Third, a model which can be applied to both monolingual and bilingual clauses allows us to cover more of the data, and consider language choice within monolingual as well as bilingual clauses. We suggest that a CS model which addresses the relationship between bilingual clauses and monolingual clauses is preferable to a model that ignores this issue by assuming that CS data are fundamentally different from monolingual data.

We evaluated four different approaches to the analysis of CS in order to determine which approach best met these three criteria. These approaches were: Poplack's (1980) CS constraints; Muysken's (2000) typological approach; a Minimalist approach (MacSwan, 2000, 2005a, 2009); and the Matrix Language Frame (MLF) model (Myers-Scotton, 2002, 2009).

First we address Poplack's (1980) model involving syntactic constraints on CS, in particular the Equivalence Constraint and Free Morpheme Constraint. The Equivalence Constraint (1980: 228) proposes that CS can only occur at a point where the order of surface morphemes from two languages follows the syntax of both those languages, while the Free Morpheme Constraint (1980:227) posits that CS cannot occur between bound morphemes, such as between a verb stem and its inflection. Criterion (1) is satisfied by Poplack's model, since the constraints are originally based on data from a corpus of spontaneous speech collected from Spanish-English bilinguals from Puerto Rico (1980: 215), and the same constraints have been tested on other bilingual speech corpora. Although the focus in Poplack's analysis is on switches and switch points as a means of testing the constraints, her classification of the syntactic categories of switched segments as either intra-sentential or extra-sentential means that the approach lends itself to a clause-based analysis and therefore satisfies Criterion (2). However, Criterion (3) is not satisfied by this model as it is not clear how it could be applied to monolingual clauses which lack switches.

Muysken's (2000) typological model of CS[1] aims to describe patterns of CS as a taxonomy rather than using the notion of constraints. Muysken suggests that CS can be categorized according to a three-part typology, where CS takes the form of either *insertion*, *alternation* or *congruent lexicalization*, all of which occur in different structural conditions and vary according to "specific bilingual settings" (2000: 3). Insertion in Muysken's typology consists of the use of one or more morphemes (including entire constituents) from one

[1] Muysken refers to intrasentential CS as code-mixing (2000: 1), but we shall retain the term CS for the purposes of this discussion.

language in a sentence which contains words from another. Alternation consists of an inter-clausal or intra-clausal switch from one language into another ("a true switch from one language to the other, involving both grammar and lexicon", 2000: 5), while maintaining congruence with the structure of both languages at the switch point. Finally, congruent lexicalization covers the phenomenon where both participating languages share structure for a given construction or part of a clause, and that shared structure can be filled with lexemes from either language or both languages (2000: 6). Muysken's CS typology has been applied to several existing corpora of bilingual production data, thus this approach satisfies Criterion (1). Muysken (2000)'s feature-based model was developed by Deuchar, Muysken and Wang (2007) into a scoring system which aimed to study CS data quantitatively at the clause level in a way that allowed the dominant CS pattern(s) for a given dataset to be identified (see Deuchar et al, 2007: 321 etc.). This approach thus satisfies Criterion (2). However, since the analysis focuses on the switch as a unit to be categorized, as in Poplack's approach, it cannot be applied to monolingual clauses and thus does not satisfy Criterion (3).

Next we discuss MacSwan's (2000, 2005) Minimalist approach to CS. the viewpoint of the Minimalist Program (MP) on CS. Authors such as MacSwan, (2005, etc.) argue that there is no need to have the concept of a "third grammar" to account for CS, since all instances of CS may be accounted for by using the grammar of monolinguals' speech, where lexically-encoded features are matched during the derivation of a sentence to produce a grammatical sentence (MacSwan, 2005: 15). It is doubtful whether the MP approach satisfies Criterion (1), since it is a generative approach concerned with speaker competence rather than production (Chomsky, 1995). While some quantitative analysis of CS using the MP has been undertaken (see Herring, Deuchar, Parafita and Moro, 2010), most of the research is concerned with grammaticality judgments and elicited data (cf. MacSwan, 2000), and it has not been extensively applied to corpora. Nevertheless, it does satisfy Criterion (2), since the CP[2] is recognized in generative grammar. It also satisfies Criterion 3 since it was originally developed for monolingual data and its proponents argue that its feature-checking approach make it easily extendable to bilingual data.

Finally we turn to the Matrix Language Frame model. Most existing studies that involve the application of the MLF model (e.g. Myers-Scotton, 1993, 2002; Schmitt, 2000; Smith, 2006; Deuchar, 2006; Deuchar and Davies, 2009;

[2] CP = 'Complementizer Phrase' (CP), which is roughly equivalent to a clause. See also footnote 2.

Table 1 Evaluation of various analytical approaches to CS

	Poplack's CS constraints	Muysken's CS typology	Minimalist Program	Matrix Language Frame model
Designed to deal with production data	✓	✓	×	✓
Can analyze individual clauses	✓	✓	✓	✓
Applies to both monolingual and bilingual clauses	×	×	✓	✓

Davies and Deuchar, 2010; Davies, 2010) analyze production data in order to identify the ways in which a bilingual's two languages interact. Indeed, we are not aware of a study that uses the MLF model to analyze anything other than production data, so it satisfies Criterion (1). Studies like Deuchar (2006), Smith (2006) and Davies (2010) have shown how the MLF model can be used to analyze CS at the clause level, thus satisfying Criterion (2). Both Smith and Davies analyze both bilingual and monolingual clauses, showing that Criterion (3) is also satisfied. Lastly, as will be described below, the MLF model allows the identification of a matrix language (ML) in each monolingual and bilingual clause. The ML can thus be seen as a dependent variable with at least two variants, depending on the language source of the ML.

Table 1 summarizes the above in that it shows which of the criteria are satisfied by the four models.

As Table 1 shows, all four models can analyze individual clauses. However, that is the only criterion that is met by all the models. The MP approach differs from the other four in that it is not designed to deal with production data. Poplack's constraints approach and Muysken's typological approach both meet two of the criteria but cannot be applied to monolingual data. The MLF model is therefore the only one that meets all of the criteria, and has been selected by us for that reason. In the next section we will outline it in more detail.

3. The Matrix Language Frame model

Myers-Scotton (2002: 8) proposes that in "classic code-switching" there is a base or Matrix Language (ML), which supplies the morphosyntactic frame for

the clause[3], and an Embedded Language (EL), which provides inserted material consisting mostly of content morphemes. The ML of a clause may be identified by applying two principles: the System Morpheme Principle (SMP) and the Morpheme Order Principle (MOP). According to the SMP, the ML sources outside late system morphemes, which are morphemes that have "grammatical relations external to their head constituent" (Myers-Scotton, 2002: 59), and thus have to look outside their maximal projection for information about their grammatical form. An example of an outside late morpheme is finite verb morphology, since verbal morphology is dependent on the subject constituent for information about its form. Example[4] (1) from the Welsh-Spanish data can be used to demonstrate how the outside late system morphemes can help identify the ML of a clause.

(1) *oedd* *o* *wedi* *gorffen* **primaria** *erbyn* *hyn ?*
be.IMP.3SG he PRT.PAST finish.NONFIN primary_school by this[5]
'Had he finished primary school by then?' (Patagonia 6)

In Example (1) all the morphemes are from Welsh, apart from the Spanish noun *primaria* ('primary school'). The subject-verb agreement shown in *oedd* (third person singular of the verb 'to be' in the imperfect tense) is an example of an outside late system morpheme, which in this case comes from Welsh. Thus, based on the subject-verb agreement morphology, the ML can be identified as Welsh[6].

---

[3] According to Myers-Scotton (2002: 54), the unit of analysis is the Complementizer Phrase (CP), which she defines as the "syntactic structure expressing the predicate-argument structure of a clause". A bilingual CP contains bilingual constituents (Myers-Scotton, 2002: 56), which may include other CPs (ie. a complex clause with embedded CPs).

[4] In all examples Welsh material is shown in *italics*, Spanish in **bold** and English in SMALL CAPS.

[5] We use the following abbreviations in glossing examples:

1SG	1st person singular inflection	M	masculine
3PL	3rd person plural inflection	NONFIN	nonfinite verb
3SG	3rd person singular inflection	PAST	past tense
DEM	demonstrative	PRES	present tense
F	feminine	PRT	nonperfective aspectual particle
INF	infinitive	PRT.PAST	perfective aspectual particle
IMP	imperfect tense	SG	singular

[6] In illustration of the way that the MLF can apply to monolingual as well as bilingual data, note that if example (1) consisted entirely of Welsh lexical items, with the Spanish *primaria* replaced by *ysgol gynradd* ('primary school' in Welsh), the conclusions about the ML would be exactly the same.

The MOP states that word order will also be sourced from the ML, with the exclusion of what Myers-Scotton (2002: 139) calls 'EL islands', which are strings of morphemes that are well-formed according to the grammar of the EL. Example (2) from the Welsh-English data illustrates the use of morpheme order to identify the ML in the clause.

(2) *mae* AMERICANS *yn* *mwy* COMMERCIAL
be.PRES.3SG PRT more
'Americans are more commercial.' (Fusser 27)

If we apply the MOP to Example (2), we can determine that the ML is Welsh given that the finite verb *mae* precedes the subject *Americans* in Welsh verb-subject order position.

The MLF model has received some criticism. Critics have targeted, among other aspects, the model's assumption that a "third grammar" is required to explain patterns in CS speech compared with monolingual speech (e.g. MacSwan, 2005a, 2005b; Gardner-Chloros and Edwards, 2004), the definitions of types of morphemes in the model (e.g. MacSwan, 2005a), and have argued that certain types of CS which adhere to the principles of the MLF may nevertheless be deemed ungrammatical in syntax judgment tasks (e.g. González-Vilbazo & López 2011). Those supporting the MLF model have responded to some of these criticisms in print (e.g. Jake, Myers-Scotton & Gross, 2005, responding directly to MacSwan, 2005a—and responded to in MacSwan, 2005b), and we direct readers to such discussions for more detailed arguments both in favor of and against the MFL model. While we do not suggest that none of the above criticisms are valid, we do highlight the fact that the MLF model has been used numerous times by different researchers with few issues (see a review in section 4 below), and propose that the MLF, while not beyond criticism, is generally convincing and appropriate to the present study, as motivated in section 2 above.

4. Previous research

In this section we will review previous research applying the MLF model to CS data as well as previous research seeking to link CS patterns analyzed by a variety of models to extra-linguistic factors.

4.1 Application of the Matrix Language Frame model

All of the previous research that has applied the MLF model to Welsh-English data has found similar patterns (Deuchar, 2006; Deuchar and Davies, 2009;

Davies and Deuchar, 2010). Deuchar (2006) analyzed a sample of 163 bilingual clauses and was able to identify Welsh as the ML in 141 clauses and found that English was the ML in only four clauses. Davies and Deuchar (2010) applied the MLF to both monolingual and bilingual finite simple clauses (n=1816), and found that Welsh was the ML in 95.43% of the total clauses. When the bilingual clauses were analyzed separately, the results showed that the ML was almost categorically Welsh (n=335/336, 99.7%). Thus, given the findings from the previous studies, we expect to find similar patterns in the choice of ML in our data.

In contrast to the research focusing on Welsh-English data, less work has been done with Spanish-English and Welsh-Spanish language pairs within the MLF framework. To the best of our knowledge, this study presents the only analysis of CS patterns in Welsh-Spanish bilingual speech. The application of the MLF model to Spanish-English CS can be found in work by Moyer, in which she investigated the "usefulness of the matrix language concept" (1995: 192). She drew attention to possible difficulties in identifying the matrix language where the two languages in question shared the same word order, and argued that predicate-argument structure, subcategorization and thematic roles should also be taken into account. Other research conducted by Smith (2006) applied the MLF model to Spanish-English data collected from Latin American immigrants in Georgia. He analyzed data from naturally occurring conversations from 56 children and adults. Smith observed that the most common type of bilingual clause in the corpus was a Spanish clause with a single English lexeme insertion, so that the ML of the clause was identified as Spanish. Smith's study also looked at the relationship between CS patterns and the following social categories: age, gender, socioeconomic status, time spent in the US, time spent in US schools. The results showed that the participants who used monolingual English and more CS were younger, female, had a higher socioeconomic status, and spent more time in the US and in US schools.

4.2 Proficiency

In her classic study of Puerto Rican Spanish-English Poplack (1980) took into account information about the participants' language background and competence. Poplack found that there was a correlation between CS types (eg. extra-sentential and intra-sentential) and bilingual ability amongst the Puerto Rican speakers. She reported that "[balanced] bilinguals produce a far greater percentage of intra-sentential switches (682/1293, or 53%) than those who are Spanish-dominant (169/542, or 31%)" (Poplack, 1980: 609). Nortier (1990) investigated the relationship between CS types and bilingual proficiency in Dutch-Moroccan Arabic speakers. Bilingual proficiency was

measured through the implementation of four methods: self-report, actual language use, judgment of monolingual proficiency by native speakers, and error analysis (Nortier, 1990: 97). Nortier compared the results from the four methods in order to determine the speakers' degree of bilingualism and linguistic competence. Nortier concluded that there was a correlation between proficiency level and the use of intrasentential and single word switches, thus supporting the findings by Poplack (1980) in her study of Puerto Rican CS in New York. Berg-Seligson (1986), on the other hand, did not find a significant difference between levels of proficiency and the types of switches produced by Hebrew-Spanish bilinguals. She argued that intrasentential CS is not a universal measure of bilingualism or an indication of balanced bilingualism, and attributed the 'universality' of the constraints studied by Poplack to the typological similarity of Spanish and English.

Bentahila and Davies (1995) studied the effects of proficiency, among other factors, on patterns of CS by comparing the speech of Moroccan Arabic-French bilingual groups that differed in age and degrees of proficiency. They found evidence that conflicted with Poplack's predictions with respect to switch types. The older, more balanced bilinguals in their study tended to switch entire clauses (25.5% of switches), whereas the younger participants who were less proficient in French preferred to switch whole noun phrases (50% of switches) rather than clauses or sentences (Bentahila and Davies, 1995: 81). However, Bentahila and Davies (1995) suggest that the difference between the behavior of the two groups can be explained not only in terms of differences in proficiency, but also in terms of their exposure to differing language contact patterns.

In a MLF analysis of CS patterns in the speech of Tembisa residents, Finlayson, Calteaux, and Myers-Scotton (1998)tested the prediction "that there is a link between proficiency in the languages involved in CS and the types of CS constituents which are produced" (1998: 415). Specifically, they found that speakers whose proficiency was higher in English produced more Embedded Language (EL) islands in English than singly occurring English lexemes. Since the use of EL islands indicates that a speaker is proficient in the syntax of the EL, this finding demonstrates that proficiency plays a role in the language chosen to produce multi-word constructions, something that will be relevant to our study on the choice of the ML.

In what Myers-Scotton has termed 'classic code-switching' (2002: 25), the participating speakers will have full proficiency in the language used as the ML, and can have any degree of proficiency in the EL. We therefore propose that in our data, the language in which the speakers are most proficient will be the ML. If they are balanced bilinguals with equal proficiency in both languages, we posit that we will either find variation in the choice of ML, or if

they only choose one ML, then the choice might be due to other factors, either structural and/or social. However, according to Myers-Scotton, bilinguals are rarely completely balanced due to a variety of reasons, such as the age and manner of acquisition of the languages, and how the speaker uses the language in daily life.

4.3 Social identity

Considerable work has been done on the relationship between language and social identity, and more specifically, ethnic identity, primarily in the field of Conversation Analysis (see Gumperz, 1982; Le Page and Tabouret-Keller, 1985; Tabouret-Keller, 1997; Sebba and Wootton, 1998; Auer, 2005; De Fina, 2007; among others). Research by Sayahi (2002), explored the relationship between identity and CS in bilingual speech in three communities: Spanish-Valencian in Spain, Spanish-English in New York, and Arabic-French in Tunisia. Through the use of questionnaires, Sayahi had participants self-report their identity and CS tendencies. According to Sayahi's findings, 68% of the Spanish-English bilinguals considered CS to be a sign of their identity (2002: 377). For the Arabic-French group, Sayahi used conversational analysis to argue that the participants switched from Arabic to French in order to reinforce their socioeconomic identity through the use of the dominant language. To our knowledge, none of the previous research has explored the link between ethnic identity and the choice of ML. Our study seeks to investigate this relationship, by employing a quantitative methodology, rather than taking a conversation analysis approach. Finlayson et al (1998: 395) argue that "language is both an index of identity and a tool of communication" and that CS provides a means not only of accommodation to interlocutors but also as a means of "projecting multiple identities for themselves" and associating with more than one social group. If we consider this from the point of view of the MLF, we posit that the choice of ML will be in the language most associated with the identity held by the speaker. Thus, a participant who has self-identified as Welsh will communicate their membership in the Welsh community through the use of Welsh predominantly as the ML. In the case of diversity of the ML, we expect a more diverse reporting of identity, as our results will show.

4.4 Social network

A social network can be described as a web of individuals who are linked to one another based on friendship, kinship, or other types of social relationships (cf. Wei, Milroy and Ching, 2000; also Milroy, 2002). When used as a tool for

analysis, it can shed light on the cultural practices of a community, including language choice and code-switching patterns (Wei et al, 2000: 157). In one of the first applications of social network analysis to a bilingual community, Gal (1978, 1979) demonstrated how the linguistic behavior of bilinguals could be influenced by the types of social contacts maintained by a speaker. Gal conducted a detailed language and social network analysis of 32 bilingual speakers of German and Hungarian in Austria. She tested the relationship between age, degree of *peasantness*, social network, and language choices. The degree of peasantness was determined through the application of eleven criteria relating to a peasant lifestyle, such as the ownership of animals. Participants' social networks were defined in terms of people the participants had interacted with either that day or over the three previous days. Gal found a correlation between language choice and peasantness of network. In particular, she reported that "The more peasants the individual has in her or his social network the greater the number of social situations in which that individual uses [Hungarian]" (Gal 1978:8). Gal concluded that "the analysis of social networks illustrates the ways in which speakers exercise control over each other's linguistic presentations of self and thereby contributes to explaining the variation between informants in their patterns of language choice" (1979: 151).

While Gal focused on language choice more than specific CS patterns, the relationship between CS patterns and social network was studied by Milroy and Wei (1995). They quantified the networks of Chinese-English bilinguals from the Tyneside community in Newcastle, UK, in order to link the proportion of Chinese contacts with observed language choices. Milroy and Wei argued that social networks could account for patterns of CS language choice better than gender, generation, and occupation. The results from their analysis revealed that the speakers with a higher proportion of Chinese contacts were more likely to use monolingual Chinese with bilingual interlocutors.

Wei et al, (2007: 141) define 'strong first-order network ties' as those members of a network who a speaker would interact with regularly, or go to directly. In our study, we examine social networks constituting five strong ties that each speaker interacts with in daily life. Through use of the social network approach, we aim to account for the language choice patterns in the three distinct bilingual communities described below.

5. The bilingual communities

The present study compares CS patterns across three bilingual communities: 1. Spanish-English bilinguals from Miami, Florida; 2. Welsh-English bilinguals

from North Wales, UK; and 3. Welsh-Spanish bilinguals from Patagonia, Argentina. In this section we present a brief description of the community members, language history, and bilingual education in each region. Through this description our aim is to portray the previous and current state of bilingualism in each community, and how its evolution has been impacted by the language medium of education.

5.1 Miami

The bilingual community under study in Miami is a relatively young community, which began in the 1960s (Gathercole, 2007), due to an influx of immigrants from Cuba after the Cuban revolution. These early immigrants were those responsible for the first Spanish-English bilingual education program that was established at the Coral Way primary school (Dade County, Miami) in 1963 (García, 2009). By the end of the 1960s the number of bilingual schools had grown to fourteen. As years passed, however, the motivation for bilingual schools diminished until only four bilingual schools remained in 1984 (García and Otheguy, 1985).

In Miami there are monolingual English speakers, bilingual Spanish-English speakers, and second language learners of English (Gathercole, 2007). Adult monolingual Spanish speakers can still be found today in Miami due to a continual influx of new Spanish speakers from Cuba and other Central and South American countries. In Miami-Dade County 59.2% of the population is Spanish-speaking[7]. According to Gathercole (2007), the most fluent bilingual speakers in Miami are the second and third generation children of Hispanic immigrants. Both languages are used by bilinguals on a daily basis, regardless of socioeconomic levels, and are visibly present on street signs, in advertising and also in literature (Gathercole, 2007).

5.2 Wales

Welsh, the indigenous language of Wales, is descended from the language spoken in Britain since *circa* 600 BC (Deuchar, 2005). English, on the other hand, was historically the language of migrants to Britain during the middle ages and later. Welsh was the official language of Wales until the 16th century, despite the upper classes increasingly turning towards English during the preceding centuries (Jones, 1993: 539), but Henry VIII's 'Acts of Union' in

[7] http://censtats.census.gov

1535–1542 firmly established English as the new official language in courts of law and public office (Jones, 1993). Deuchar (2005: 621) describes these laws as the "final stage of the linguistic colonization of Wales". The repercussions of these measures influenced many aspects of Welsh society for many centuries, particularly in the case of education, where English was the *de facto* language of tuition in schools until the 20th century—indeed, it was the unofficial policy of some schools during the 19th century to punish pupils for speaking Welsh in the classroom (Jones, 1993: 548).

About 21% of the current Welsh population speaks Welsh, according to the 2001 UK census, although proportions of Welsh speakers today vary from area to area, from a high proportion in Gwynedd in the north (69%) to much lower proportions in other regions, e.g. 9.3% in Monmouthshire and 9.5% in Blaenau Gwent[8]. A large number of Welsh speakers are native or native-like in both Welsh and English (Thomas and Gathercole, 2005), and no speakers are thought to be monolingual in Welsh (except perhaps very young children).

Today, a large proportion of the school-age population in Northwest Wales attends school through the medium of Welsh[9]. The first Welsh-medium primary school was opened in 1947 and the first secondary school followed in 1956 (Baker and Prys Jones, 1998). The 1993 Welsh Language Act[10] declared Welsh and English to be equal in many aspects of public life, and since 1999 Welsh has been a compulsory subject in primary and secondary schools. Thus, children are now able to learn Welsh at school, at home, or both. Welsh is also visible in daily life: on signs, in literature and the media, on Welsh television, and in official government documents. There is also support for Welsh in cultural activities: for example, Welsh-speakers have the opportunity to participate in Welsh choirs or music groups, and often take part in Welsh festivals called *eisteddfods*.

5.3 Patagonia

The Welsh community in Patagonia was first established in the 19th Century as a response to the increasing dominance of English and the Anglican Church in Wales. Over the course of fifty years from 1865 onwards, 3000 Welsh

[8] These results are taken from the following document: http://www.byig-wlb.org.uk/english/publications/publications/332.doc or http://tinyurl.com/censuscomparison

[9] Lewis (2008) reports that, in 2007, 30.5% of primary schools in Wales taught through Welsh, and 15.4% of secondary school pupils in Wales were being taught Welsh as a first language.

[10] See e.g. http://www.legislation.gov.uk/ukpga/1993/38/contents

speakers emigrated to Argentina and settled in the Chubut Valley (Rawson, Trelew, Gaiman, and Dolavon) and along the Cordillera de los Andes (Trevelin, Esquel), having been allocated land by the Argentinean government (Johnson, 2009). The original motivation for establishing the colony was to create a Welsh community isolated from other settlements, where the Welsh people could preserve their language, culture, and religion. Initially, the colony was self-sustaining, and had its own government and education system, although the administrative practices of the government were conducted in Spanish. In time, however, it became a legal requirement to attend school through the medium of Spanish and the Welsh community began to integrate into the wider Spanish-speaking community. Currently there is a recently established bilingual Welsh-Spanish medium primary school in Trelew, and students may study Welsh as part of their curriculum in schools in the Gaiman region. (Johnson, 2009).

During the period between 1920 and 1965, Spanish became the lingua franca of the Chabut province (Gutiérrez and Jones, 2006). As the Welsh lost their status in society due to economic developments and the end of immigration, Welsh was used mostly by Welsh-speaking families and in certain religious circles (Johnson, 2009). There has been, however, an improvement in the attitude toward the Welsh language since the 1990s, and a revival of the *eisteddfod*. As in the case of the *eisteddfods* in Wales, these Welsh festivals are popular annual competitions for choirs, folk dancing, poetry, and other performances. Often Welsh performers from Wales travel to Patagonia on vacation and to participate in the local festivals.

6. The Focus of our Study

As outlined in the introduction, we are interested in how clause-based CS patterns may be related to both structural and extra-linguistic factors in the three communities. This question may be formulated in terms of specific predictions which we tested on our data. Our first prediction addresses the possible role of structural, language-internal factors in CS patterns. Chan (2009) has suggested that there may be a universal tendency to select only one ML unless the two languages have similar word order. When a speaker is code-switching between a SVO and SOV language, the most optimal option is to activate either word order, rather than the less optimal options of activating both or neither. The choice between the two word order options, however, is governed by sociolinguistic factors (2009: 193). Our three corpora vary in that one (Spanish-English) involves languages with a similar SVO word order whereas

the other two (Welsh-English and Welsh-Spanish) involve languages with contrasting orders, VSO and SVO. If Chan is right then we may expect a relatively uniform choice of ML in the Welsh-English and Welsh-Spanish data, but a variable choice in the Spanish-English data.

Our second prediction concerns extra-linguistic factors which affect bilingual speech patterns, and thus may specifically predict ML choice. This applies to proficiency, ethnic identity, and social network. Our predictions in relation to these factors as are follows:

1. Proficiency: we expect to find a high level of proficiency in the language(s) selected as ML. This is in line with Myers-Scotton's (2002: 25) claims that high proficiency is needed in a language for it to be used as an ML, and is supported by the correlations that have been established between proficiency levels and CS types (Poplack, 1980; Nortier, 1990; Finlayson et al, 1998).
2. Identity: where there is homogeneity, the language associated with that identity will also be the most common ML. This prediction is based on the finding that CS can be seen as a tool for projecting or reinforcing identity (cf. Finlayson et al, 1998, Sayahi, 2002).
3. Social network: where there is homogeneity, the most common language of the social network will also be the most common ML. This is based on evidence that patterns of language choice can be influenced by social contacts (cf. Gal, 1978, 1979; Wei et al, 2007).

7. Method

7.1 Data collection

In order to maintain consistency and comparability, a similar method was used for participant recruitment and data collection for all three corpora. Potential participants were recruited through a variety of means, such as media announcements, recruitment letters, posters, and the "friend of a friend" approach (see Milroy, 1987). 151 Welsh-English participants from Wales, 85 Spanish-English participants from Miami, and 92 Welsh-Spanish participants from Patagonia were recorded having natural conversations in pairs or groups of three for approximately half an hour using either a Zoom or Marantz portable recording device. The participants were able to choose their own bilingual recording partner and often chose a friend or family member. They could speak freely about a topic or topics of their choice. In order to minimize the Observer's Paradox (Labov, 1972), the investigator was not present for the

duration of the conversation. The digital audio was recorded then later transcribed using the CHAT system (MacWhinney, 2000).

After the conversations were recorded, the participants were administered a questionnaire in order to elicit background information. Questionnaires relating to all three corpora were highly similar in design, although they were adapted slightly for each community because of differences relating to the linguistic and cultural context. Of particular interest to the present study are the questions that address language proficiency, identity, and the languages of the social network. The participants were asked to give a self-assessment of their proficiency in each of the languages based on a four point rating scale: i) basic, ii) some words and expressions, iii) fairly confident, iv) confident. For the question of identity, the participants were given a list of three social identities plus an 'other' option of listing an additional identity or even a mixed identity. In Wales, participants could choose between the following: Welsh, British, English and other. In Miami, the options were Cuban, American, Puerto Rican and other. In Patagonia, the choices were Argentinean, Patagonian, Welsh and other. The questionnaire also included an assessment of the main language or languages of the participant's social network. Participants were asked to list the five people with whom they conversed most and to indicate which language they tended to use while speaking to them (eg. in Miami: Spanish, English or both).

7.2 Data analysis

Three transcripts were chosen from each of the corpora for the MLF analysis. We ensured that the speech in the transcripts was authentic by comparing the language(s) used in the transcript with language information from the questionnaires. Specifically, we verified that the language(s) spoken in the recorded conversation was the language(s) that the participant indicated that they usually use when speaking to their recording partner. This information was obtained from the social network question.

Once the transcripts were selected, all of the monolingual and bilingual clauses from the transcripts were extracted either manually or automatically with the Computerized Language Analysis (CLAN) program (MacWhinney, 2000). 1860 clauses were extracted from the Wales corpus, 2614 from the Miami corpus, and 1548 from the Patagonia corpus. Complex clauses were divided into component simple clauses which were separately listed[11].

[11] Simple clauses constitute one unit of analysis, but complex clauses were divided into their constituent 'clausal units'. For example, in the sentence below, which is a complex clause

All simple clauses were then coded as either monolingual or bilingual and finite or non-finite, depending on whether they contained a finite verb. For the analysis of the matrix language we used only the finite clauses, so that sentence fragments and non-finite clauses were excluded from the analysis. Each finite clause was then coded as having either a Welsh, English or Spanish ML. Example (3) below illustrates the application of the MLF to a Spanish-English bilingual finite simple clause from the Miami corpus.

(3) **ella** **es** AN ACCOUNTANT
she be.PRES.3SG
'She is an accountant.' (Sastre 1)

Since word order is the same in Spanish and English, the MOP does not help us identify the ML in the clause. However, we can use the SMP. The subject-verb agreement shown in **es** (third person singular of the verb "to be" in the present tense) is an example of an outside late system morpheme, which in this case comes from Spanish. Thus, based on the subject-verb agreement morphology, the ML can be identified as Spanish.

Extra-linguistic factors were calculated through the analysis of the questionnaire responses for every participant in each corpus. All of the data from the questionnaires were entered into a spreadsheet and then analyzed to determine relative bilingual proficiency, social identity, and the social network languages. In order to determine the main social network language for each community, first we calculated the mean scores per participant and then we calculated the mean score overall for each group (Miami, Wales, Patagonia). Although our analysis of these factors uses the data from all participants in each corpus, the social characteristics of those speakers providing the linguistic data sample are shown in the Appendix.

8. Results

8.1 Matrix Language distribution

Before we present the results from the analysis of the MLF analysis and the analysis of the extra-linguistic factors, it is necessary to give a description of the clauses extracted from the three corpora, with respect to their linguality

---

consisting of CP2 within CP1, we analyzed CP1 (I thought) as one unit—without taking account of the embedded CP2—and CP2 (it was her iPod) as a second, discrete, unit of analysis.

[CP1 I thought [CP2 it was her iPod.]]

(monolingual or bilingual status). This will illustrate the proportion of code-switched clauses vs. monolingual clauses for each group. Figure 1 gives the distribution of the monolingual and bilingual finite simple clauses in each dataset. In the Miami dataset, 94% (n=2460) were monolingual finite simple clauses and 6% (n=151) were bilingual finite simple clauses. For Wales, 81% (n=1515) were monolingual and 19% (n=345) were bilingual clauses. The Wales dataset has the highest percentage of bilingual clauses out of the three groups. In the Patagonia dataset, 97% (n=1503) were monolingual finite simple clauses and only 3% (n=45) were bilingual.

Figure 2 presents the distribution of the ML for the monolingual finite simple clauses. We include the analysis of the monolingual data in order to give a fuller picture of the ML choice in the datasets. For Miami, out of the total number of monolingual finite clauses, 51% (n=1250) had a Spanish ML and 49% (n=1210) had English as the ML. For Wales, 95% (n=1433) were identified as having a Welsh ML and 5% (n=82) had English as the ML. Finally for Patagonia, a Spanish ML was identified in 19% (n=289) of the monolingual clauses and Welsh was identified in the remaining 81% (n=1214).

The distribution of the ML for the bilingual finite simple clauses is shown in Fig. 3. 100% (n=345) of the bilingual clauses from the Wales dataset had

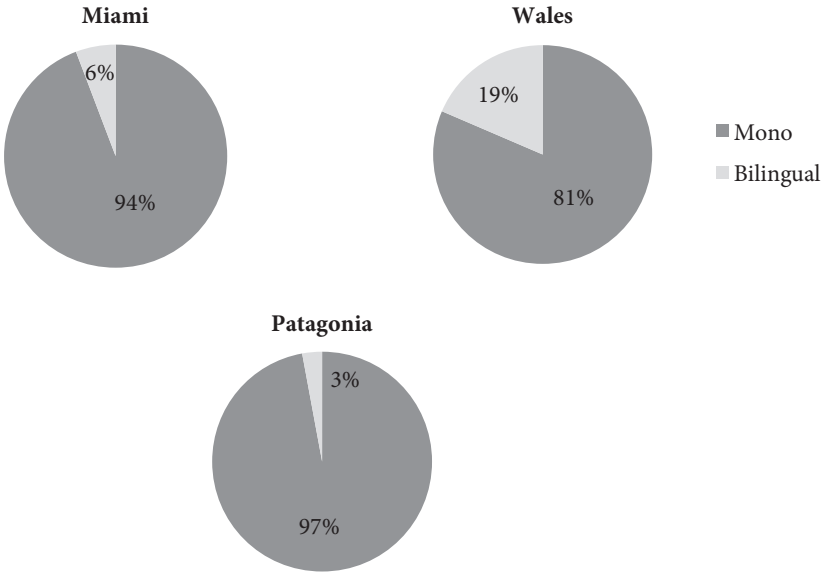

Figure 1: Overall distribution of monolingual and bilingual clauses.

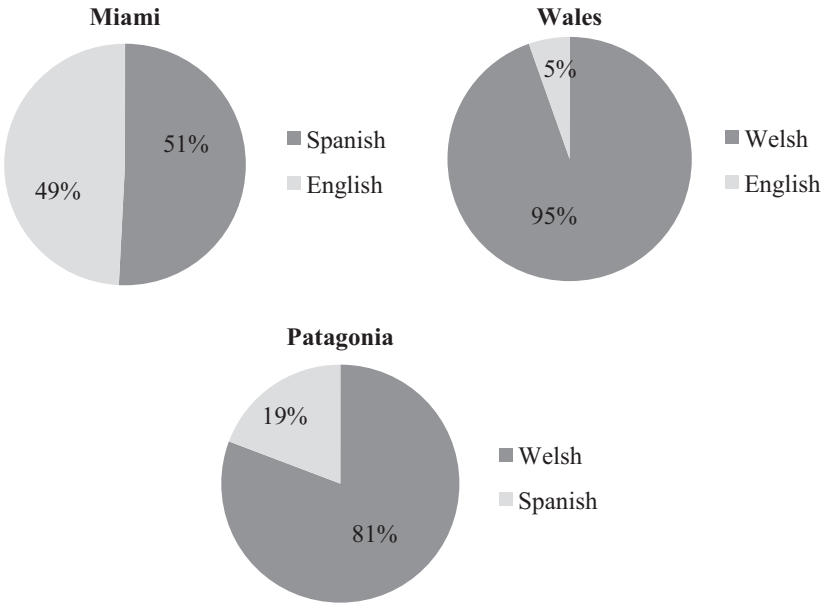

Figure 2: Matrix Language distribution of monolingual clauses.

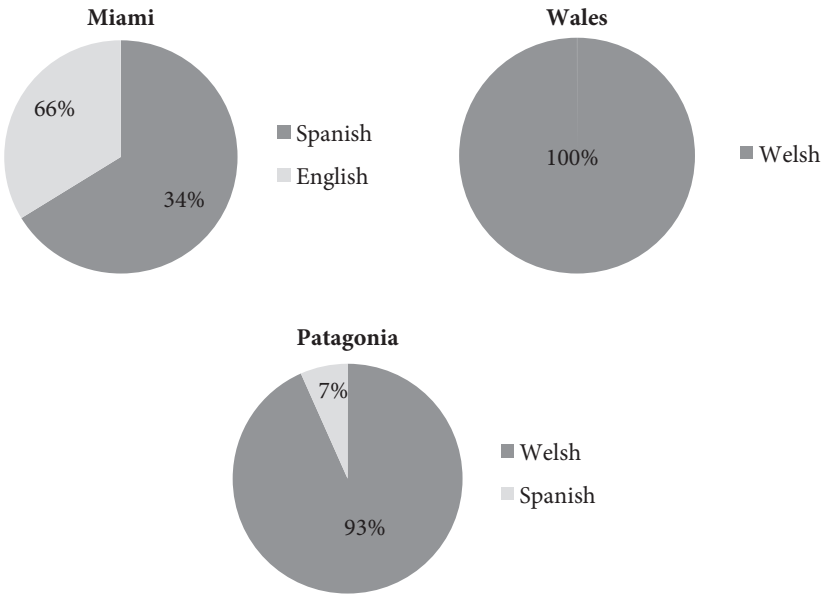

Figure 3: Matrix Language distribution of bilingual clauses.

Welsh as the ML. The Patagonia data showed a similar trend as 93% (n=42) of the bilingual clauses had Welsh as the ML. In contrast, the Miami data showed more variability as 66% (n=100) of the bilingual clauses had a Spanish ML and the remaining 34% (n=51) were identified as having English as the ML.

To summarize the results from the MLF analysis, with respect to structural differences in the language pairs, we found that there was uniformity in the choice of ML when the language pair has a different word order (VSO-SVO), which is the case with Welsh-English and Spanish-Welsh language pairs. We also found that there was variability when the language pair has a similar word order, as with the Spanish-English data (SVO-SVO). The structure of the two languages involved in the code-switching, however, does not account for the actual language choice of the ML (Chan, 2009). Therefore, we turn to extra-linguistic factors to explain our findings.

8.2 Proficiency

Figures 4, 5, and 6 show the relative bilingual proficiency overall for the three communities. The Miami participants were balanced bilinguals, with 73% (n=62) of the participants having equal proficiency in Spanish and English (Fig. 4). The majority of the Wales participants (63%, n=94) were also

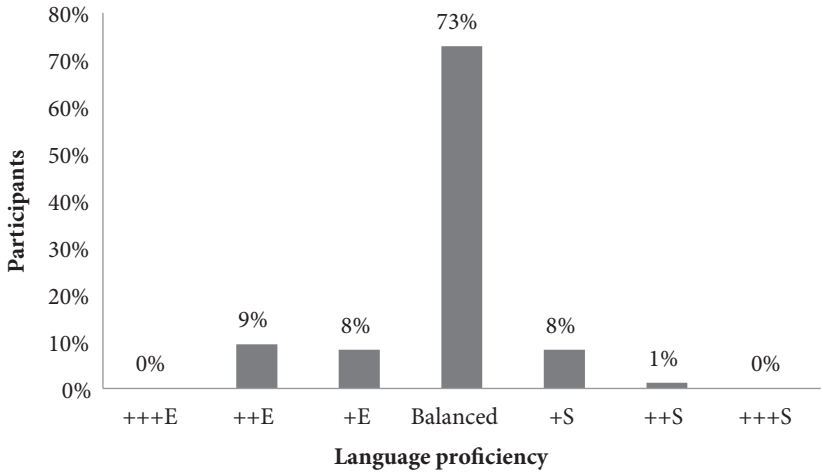

Figure 4: Relative balanced proficiency in Miami. E indicates a higher English proficiency; S indicates a higher Spanish proficiency.

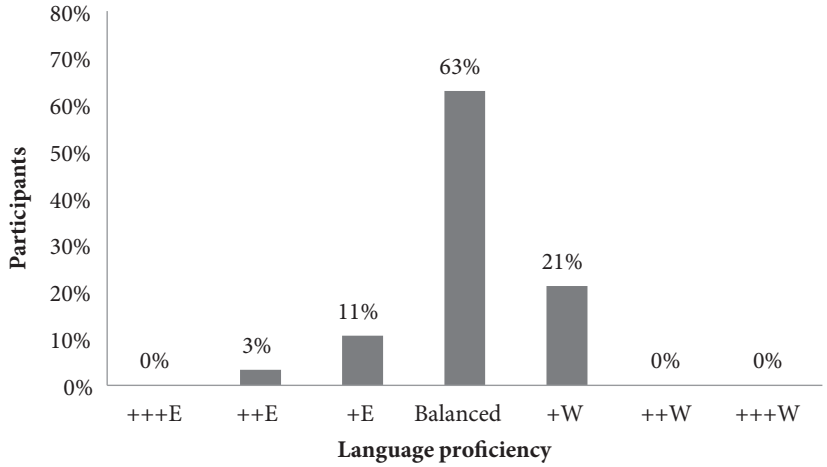

Figure 5: Relative balanced proficiency in Wales. E indicates a higher English proficiency; W indicates a higher Welsh proficiency.

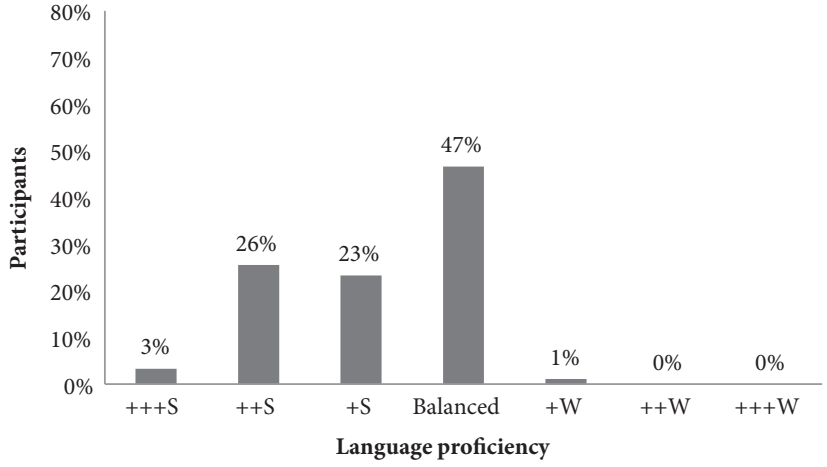

Figure 6: Relative balanced proficiency in Patagonia. S indicates a higher Spanish proficiency; W indicates a higher Welsh proficiency.

balanced bilinguals, with equal proficiency in both Welsh and English. 21% (n=32) of the Wales participants were more confident in Welsh than in English (Fig. 5). The participants from Patagonia, on the other hand, were divided almost equally between balanced bilinguals and those who were more proficient in Spanish.

8.3 Identity

Notable differences can be seen between the three communities with respect to the self-reported social identity (Fig. 7). 90% (n=136) of the Wales participants self-identified as Welsh. Other choices were British (3%), English (3%), or Other (4%). The majority of the Patagonia participants (62%, n=55) reported that they were Argentinean and 20% (n=18) self-identified as Patagonian. Only 6% (n=5) reported their identity as Welsh. 12% (n=11) chose both Patagonian and Argentinean as their response, indicating a dual identity. The results for the Miami participants were more varied. Approximately half of the participants elected to indicate an identity that fell into the Other category (46%, n=41), which included Venezuelan, Nicaraguan, and Dominican. Cuban-American among other mixed identities represented 20% of the total responses. The remaining participants self-identified as American (30%, n=27) and Cuban (20%, n=18).

8.4 Social network

The results of the social network analysis are illustrated in Fig. 8. A score of 3 indicates that the main language of speakers' social network in the majority

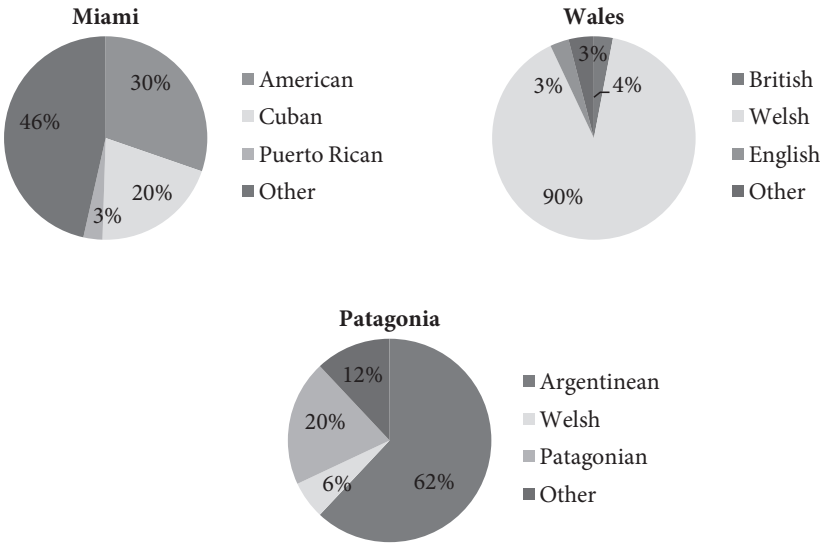

Figure 7: Social identity in Miami, Wales and Patagonia.

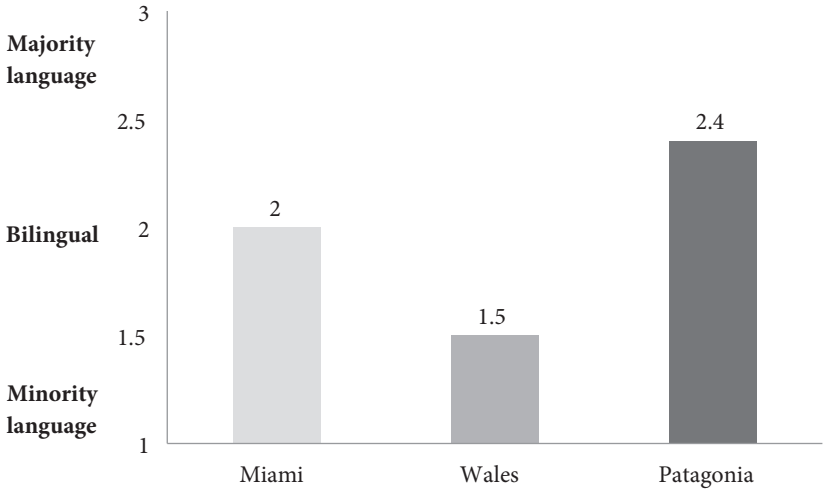

Figure 8: Social network mean scores for Miami, Wales and Patagonia. 3 indicates a social network in the majority language; 2 indicates a bilingual network; 1 indicates a social network in the minority language.

language of the respective country, which is English in the case of Wales and Miami, and Spanish in Patagonia. A score of 2 indicates that the speakers use both languages in their social networks. A score of 1 signifies that the minority language is the main language of the social network (Welsh in Wales and Patagonia, and Spanish in Miami). The participants from Miami generally have a bilingual network score (M = 2), the participants from Wales tend to have a more Welsh-speaking social network (M = 1.5), and the participants from Patagonia have a more Spanish-speaking social network (M = 2.4).

9. Discussion

The ML distribution in both monolingual and bilingual clauses shows considerable diversity (with both English and Spanish as ML) in the Miami data compared with considerable uniformity in the Welsh-English and Welsh-Spanish data. In the data collected both in Wales and Patagonia, Welsh is overwhelmingly the most frequent ML, almost to the exclusion of the other language, including bilingual clauses. These results are in line with our prediction that more uniformity of the ML would be found where the two languages in contact have different word orders, and this is the case with the VSO language (Welsh) in contact with an SVO language (English or Spanish).

We now consider the influence of the extra-linguistic factors, starting with (self-assessed) proficiency. In Miami and Wales we have seen that balanced bilingualism is a common feature of the speakers. Myers-Scotton (2002: 25) argues that full proficiency in the language of the ML is needed for it to be used as an ML in classis CS, and if this is the case then it looks as though speakers in both communities should generally have a high enough level of proficiency to be able to choose either of their two languages as ML. In Miami it seems that they do choose freely, but not in Wales, where Welsh as an ML is strongly preferred over English. But we have already suggested that structural factors favor uniformity, resulting in the choice of one language rather than the other, though it remains for us to account for why the ML chosen should be Welsh rather than English.

In Patagonia we have seen that fewer than half of speakers are balanced bilinguals, and the rest have mostly superior proficiency in their Spanish in contrast to English. However, for reasons to be discussed below, the Patagonia data analyzed for this paper turn out not to be fully representative of the community as a whole, at least partly because the average age of the speakers analyzed was 72. Speakers over 50 in Patagonia in fact show a greater degree of balanced bilingualism than the rest of the population, as can be seen in Figure 9 below. If this fact is taken into account then the speakers' choice of Welsh as ML is not so unexpected, but we still need to explain why Welsh is chosen in preference to Spanish, in which they are at least as proficient. The relationship between age and proficiency levels, and subsequently ML choice,

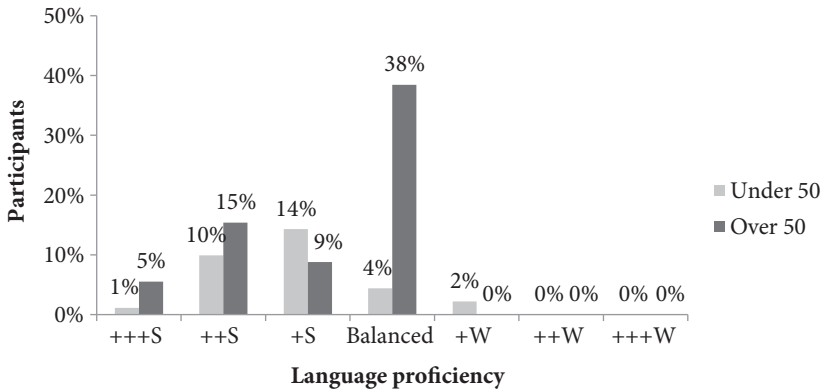

Figure 9: Relative balanced proficiency in the Patagonia corpus divided according to age: over 50 years or under 50 years. S indicates a higher proficiency in Spanish; W indicates a higher proficiency in Welsh.

should be explored further in future research, and could be a sign of an occurrence of language change within the Patagonia community, where, for example, Welsh as a first language is dying out, and it is becoming more common as a culturally-marked second language.

Our results on identity showed considerable homogeneity in Wales, where most participants indicated Welsh as their identity, but considerable diversity in Miami and Patagonia, with the most popular identity being American in Miami and Argentinean in Patagonia. In Wales and Miami we can argue that the uniformity vs. diversity of self-identity may be related to the corresponding uniformity/diversity of the ML. However, we cannot argue the same in Patagonia, where our speakers choose mostly Welsh ML despite selecting Argentinean identity, which one would expect to be associated more with Spanish. In future research, one possibility would be to give a mixed identity as a specific option in the questionnaire, such as a Patagonia-Argentinean identity, rather than expect participants to indicate this in the 'other' option. Several of the Miami participants (20%) listed two or more nationalities as their identity, but far fewer participants from Patagonia gave a combined identity as their response (12%). Again, the special case of Patagonia will be discussed below.

Finally, Patagonia is once again the odd one out in relation to the implications of the social network scores. The tendency of Welsh-English speakers in Wales to have a mainly Welsh-speaking social network fulfils our prediction that the most common language of the social network will also be the most common ML. Likewise, the tendency of Spanish-English speakers in Miami to have a diverse or bilingual network leads us to expect diverse MLs, which again supports the same prediction. But the preference for Welsh ML on the part of our Welsh-Spanish speakers in Patagonia does not appear to be predicted by the fact that their networks tend to be more Spanish-speaking than Welsh-speaking. It could be that their strong network ties in daily life are predominantly Spanish-speaking, and that the use of Welsh is limited to only a restricted number of people, including the interlocutor selected for the recording. A more detailed and extensive social network analysis, including a larger number of ties, is needed to thoroughly investigate this factor. The approach could involve the analysis of between twenty and fifty ties (cf. Wei et al, 2007) and include both strong and weak associations.

As we have suggested, our results from Patagonia require some explanation. Although they do meet the prediction that uniformity of the ML will be the norm because of the structural difference between Welsh and Spanish, they do not meet our predictions regarding choice of ML from the point of view of proficiency, identity, or social network. Therefore, an alternative explanation

should be sought in future research in order to account for the speech patterns found in the Patagonia data. A reason for this may be that that unlike Welsh-English speakers in Wales and Spanish-English speakers in Miami, the Welsh-Spanish speakers do not really form a community in which both languages are part of the majority of speakers' repertoire. While the proportion of Welsh-speakers in many parts of Wales is at least 60% or higher, and the proportion of Spanish speakers in Miami was estimated to be 67%[12] in 2000 only a small minority of the population in Patagonia speaks Welsh (approximately between 250-500 people), and Welsh is rarely heard casually on the street, for example. This may mean that while the effect of structural factors on Welsh-Spanish speakers' conversation is still relevant, other social factors not directly addressed in this paper may play a role. For example, people in the community may use Welsh only with a limited number of people, and then not typically in a code-switching mode. This is supported by the fact that, of the three corpora, the Patagonia data had the lowest proportion of bilingual clauses (the type with intrasentential CS). We can only speculate at this point as to why we find mostly monolingual Welsh clauses in our data. It is possible that the bilinguals in this community share a purist ideology towards the Welsh language, and thus tend to refrain from speaking in a code-switching mode. However, if this were the case, we would assume that the speakers would self-identify as Patagonian or even Welsh. In our data, as shown in our analysis of identity, the majority identified as Argentinean, only 20% identified as Patagonian, and 12% chose both Argentinean and Patagonian options. Another explanation may be that speakers reproduce the patterns of language contact to which they are exposed, and exposure to monolingual rather than bilingual clauses will tend to lead to production of the same. There may also be other factors influencing the speech of the Spanish-Welsh bilinguals that should be explored in future research.

10. Conclusions

The main question addressed in this paper was how CS patterns were related to structural and extra-linguistic factors in three bilingual communities. In order to answer this question, we first analyzed CS patterns by applying the MLF model to nine transcripts of naturalistic conversation held between bilinguals in Miami, Wales and Patagonia. We found uniformity in the choice of Welsh as the ML in the Wales and Patagonia data, whereas in the Miami

[12] See http://www.census.gov/prod/2003pubs/c2kbr-29.pdf

data there was variability in the choice of ML. We suggested that this uniformity vs. variability contrast might be due to structural linguistic factors, given that there may be a universal tendency to select one ML when the two languages involved have different word order. We also investigated the effects of extra-linguistic factors on the distribution of ML in the three corpora, specifically proficiency, identity and social networks. In the case of Miami, the speakers had balanced proficiency in their two languages and bilingual social networks. There was also diversity in their self-reported identities. All three of these factors may be related to the diversity we find in the distribution of the ML in Miami. For Wales, on the other hand, we found a relationship between uniformity in the distribution of the ML and uniformity in the speakers' identities and social networks. The results for Patagonia were not as clear. Although the structural difference between Welsh and Spanish predicted uniformity in the ML, the choice of ML could not be accounted for through the study of the factors included in this study. We suggest that unlike in Miami and Wales, the bilingual speakers do not form part of a linguistic community where they use both languages regularly. Instead they may use Welsh unilingually and only with a small group of people.

Acknowledgements

The support of the Arts and Humanities Research Council (AHRC), the Economic and Social Research Council (ESRC), the Higher Education Funding Council for Wales and the Welsh Assembly Government is gratefully acknowledged. The Welsh data were originally collected and transcribed by Marika Fusser, Elen Robert, Peredur Davies, Jonathan Stammers, Siân Lloyd-Williams and Margaret Deuchar as part of the project "Code-switching and Convergence in Welsh", awarded by the AHRC to Margaret Deuchar. They are available via TalkBank (http://www.talkbank.org) in the BilingBank archive. The work presented in this paper was part of the programme of the ESRC Centre for Research on Bilingualism in Theory and Practice at Bangor University. We are indebted to Fraibet Aveledo, Alberto Rosignoli, and all those who helped with the transcription of the three corpora used in this study.

References

Auer, Peter. 2005. A postscript: code-switching and social identity. *Journal of Pragmatics* 37: 403-410.

Baker, Colin and Sylvia Prys Jones. 1997. *Encyclopaedia of Bilingualism and Bilingual Education*. Clevedon: Multilingual Matters.

Bentahila, Abdelâli and Eirlys Davies. 1995. Patterns of code-switching and patterns of language contact. *Lingua* 96: 75-93.

Berk-Seligson, Susan. 1986. Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15: 313-348.

Chan, Brian. 2009. Code-switching between typologically distinct languages. In Barbara Bullock and Almeida Jaqueline Toribio (eds.), *The Cambridge Handbook of Linguistic Code-switching*, 182-198. Cambridge: Cambridge University Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Davies, Peredur. 2010. *Identifying word-order convergence in the speech of Welsh-English bilinguals*. PhD dissertation, Bangor University.

Davies, Peredur and Margaret Deuchar. 2010. Using the Matrix Language Frame model to identify word order convergence in Welsh-English bilingual speech. In Anne Breitbarth, Christopher Lucas, Sheila Watts, and David Willis (eds.), *Continuity and Change in Grammar*, 77-96. Amsterdam: John Benjamins.

De, Fina, Anna. 2007. Code-switching and the construction of ethnic identity in a community of practice. *Language in Society* 36: 371-392.

Deuchar, Margaret. 2005. Minority language survival in Northwest Wales: an introduction. *ISB4: Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press.

Deuchar, Margaret. 2006. Welsh-English codeswitching and the Matrix Language Frame model. *Lingua* 116: 1986-2011.

Deuchar, Margaret and Peredur Davies. 2009. Code-switching and the future of Welsh. *International Journal of the Sociology of Language* 195: 15–38.

Deuchar, Margaret, Pieter Muysken, and Sung-Lan Wang. 2007. Structured Variation in Codeswitching: Towards an Empirically Based Typology of Bilingual Speech Patterns. *The International Journal of Bilingual Education and Bilingualism* 10: 298-340.

Finlayson, Rosalie, Karen Calteaux, and Carol Myers-Scotton. 1998. Orderly mixing and accommodation in South African codeswitching. *Journal of Sociolinguistics* 2(3): 395-420.

Gal, Susan. 1978. Peasant men can't get wives: language change and sex roles in a bilingual community. *Language in Society* 7:1-16.

Gal, Susan. 1979. *Language Shift: Social Determinants of Linguistic Change in Bilingual Austria*. New York: Academic Press.

García, Ofelia and R. Othegy. 1985. The masters of survival send their children to school: Bilingual education in the ethnic schools of Miami. *Bilingual Review/Revista Bilingüe* 12: 3-19.

García, Ofelia. 2009. El uso del español en la enseñanza. La educación bilingüe. In Humberto López Morales (ed), *Enciclopedia del español en los Estados Unidos*, 417-422. Instituto Cervantes, Santillana.

Gardner-Chloros, Penelope. 2009. Sociolinguistic factors in code-switching. In Barbara Bullock and Almeida Jaqueline Toribio (eds.), *The Cambridge Handbook of Linguistic Code-switching*, 97-113. Cambridge: Cambridge University Press.

Gardner-Chloros, Penelope and Malcolm Edwards. 2004. Assumptions behind grammatical approaches to code-switching: When the blueprint is a red herring. *Transactions of the Philological Society* 102: 1433-1452.

Gathercole, Virginia. 2007. Miami and North Wales, So Far and Yet so Near: A Constructivist Account of Morphosyntactic Development in Bilingual Children. *The International Journal of Bilingual Education and Bilingualism* 10: 224-247.

González-Vilbazo, Kay and Luis López. 2011. Some properties of light verbs in code-switching. *Lingua* 121(5): 832-850.

Gumperz, John. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.
Gutiérrez, Graciela and Vilma Nanci Jones. 2006. Mantenimiento y cambio del galés en el valle del Río Chabut. In *Los Galeses en la Patagonia III*, 191-216. Puerto Madryn: Asociación Punta Cuevas.
Herring, Jon, Margaret Deuchar, María del Carmen Parafita Couto, and Mónica Moro Quintanilla. 2010. 'I saw the madre': evaluating predictions about codeswitched determiner-noun sequences using Spanish-English and Welsh-English data. *International Journal of Bilingual Education and Bilingualism* 13 (5): 553-573.
Jake, Janice, Carol Myers-Scotton, and Steven Gross. 2002. Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism, Language and Cognition* 5: 69-91.
Johnson, Ian. 2009. How green is their valley? Subjective vitality of Welsh language and culture in Chubut province, Argentina. *International Journal of the Sociology of Language* 195: 141-171.
Jones, Robert O. 1993. The sociolinguistics of Welsh. In Martin J. Ball with James Fife (eds.), *The Celtic Languages*, 536–605. London: Routledge.
Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1: 97-120.
Le Page, Robert and Andrée Tabouret-Keller. 1985. *Acts of Identity: Creole-based approaches to ethnicity and language*. Cambridge: Cambridge University Press.
Lewis, W. Gwyn. 2008. Current challenges in bilingual education in Wales. *AILA Review* 21: 69-86.
MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: Evidence from codeswitching. *Bilingualism: Language and Cognition 3* (1): 37–54.
MacSwan, Jeff. 2005a. Remarks on Jake, Myers-Scotton and Gross's response: There is no "Matrix Language". Bilingualism: Language and Cognition 8: 277–284.
MacSwan, Jeff. 2005b. Code-switching and generative grammar: A critique of the MLF model and some remarks on 'modified minimalism'. Bilingualism: Language and Cognition 8: 1-22.
MacSwan, Jeff. 2009. Generative approaches to code-switching. In Barbara Bullock and Almeida Jaqueline Toribio (eds.), The Cambridge Handbook of Linguistic Code-switching, 309-335. Cambridge: Cambridge University Press.
MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Milroy, Lesley. 1987. *Language and Social Networks*. Oxford: Blackwell.
Milroy, Lesley. 2002. Social networks. In J. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.), *Handbook of Language Variation and Change*, 549-572. Oxford: Blackwell Publishing.
Milroy, Lesley and Lee Wei. 1995. A social network approach to code-switching. In Lesley Milroy and Pieter Muysken (eds.), *One speaker, Two Languages*, 136-157. Cambridge: Cambridge University Press.
Moyer, Melissa. 1995. On defining matrix language in sentential code-switching: a syntactic approach. *Proceedings of the Summer School of code-switching and language contact of The Netherlands*: 192-204.
Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
Myers-Scotton, Carol. 1993. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
Myers-Scotton, Carol. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford and New York: Oxford University Press.

Myers-Scotton, Carol. 2009. A universal model of code-switching and bilingual language processing and production. In Barbara Bullock and Almeida Jaqueline Toribio (eds.), *The Cambridge Handbook of Linguistic Code-switching*, 336-357. Cambridge: Cambridge University Press.

Nortier, Jacomine. 1990. *Dutch-Moroccan Arabic Code Switching among Moroccans in the Netherlands*. Dordrecht: Foris Publications.

Poplack, Shana. 1980. Sometimes I'll start a sentence in English y termino en Español: towards a typology of code-switching. *Linguistics* 18: 581–618.

Poplack, Shana. 1988. Contrasting patterns of codeswitching in two communities. In Monica Heller (ed.), *Codeswitching: anthropological and sociolinguistic perspectives*, 215-244. Berlin: Walter de Gruyter.

Sayahi, Lotfi. 2002. Social identity and code-switching in bilingual contexts: a comparative approach. In Leonard Ashley and Wayne Finke (eds.), *Language and Identity: Selected Papers of the International Conference*, *October 2-5*, *2002*, 373-381. East Rockaway, NY: Cummings & Hathaway Publishers.

Schmitt, Elena. 2000. Overt and covert codeswitching in immigrant children from Russia. *International Journal of Bilingualism* 4: 9–28.

Sebba, Mark and Tony Wootton. 1998. We, they and identity: Sequential versus identity-related explanation in code-switching. In Peter Auer (ed.), *Code-Switching in Conversation*, 262-289. London: Routledge.

Smith, Daniel. 2006. Thresholds leading to shift: Spanish/English codeswitching and convergence in Georgia, U.S.A. *International Journal of Bilingualism* 10 (2): 207-240.

Tabouret-Keller, Andrée. 1997. Language and Identity. In Florian Coulmas (ed.), *The Handbook of Sociolinguistics*, 315-326. Oxford: Blackwell Publishers.

Thomas, Enlli and Virginia Gathercole. 2005. Minority language survival: Obsolescence or survival for Welsh in the face of English dominance? *ISB4: Proceedings of the 4th International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press.

Wei, Li, Lesley Milroy and Pong Sin Ching. 2007. A two-step sociolinguistic analysis of code-switching and language choice: the example of a bilingual Chinese community in Britain. In Wei Li (ed.), *The Bilingualism Reader*, 2nd edition, 139-158. London and New York: Routledge.

Appendix

Table 2 Individual speaker information

Participant ID	Age	Gender	More Proficient Language	Identity	Social Network Score
Miami					
AME	26	F	Balanced	Venezuelan	1.6
CAR	21	F	Balanced	American	2
KEV	57	M	Balanced	Cuban	2
PAI	33	F	Balanced	Cuban-American	2.2
SAR	34	F	English	Cuban-American	2
SOF	44	F	Balanced	Cuban	2.4
Wales					
AMR	36	F	Balanced	Welsh	1.8
ANT	52	F	Balanced	Welsh	1.8
DAN	25	M	Welsh	Welsh	1
HEC	23	M	Balanced	Welsh	1.4
LIS	20	F	Balanced	Welsh	1
MAB	19	F	Balanced	Welsh	1.4
Patagonia					
CHT	66	F	Balanced	Argentinean	1.8
AVR	82	F	Balanced	Argentinean	2.2
CRL	54	F	Spanish	Welsh	2.4
SAR	96	F	Balanced	Argentinean	1.8
BLA	65	F	Balanced	Patagonian	1.8
TER	69	F	Balanced	Argentinean	2.2