

# FACTORS DETERMINING CODE-SWITCHING PATTERNS IN SPANISH-ENGLISH AND WELSH-ENGLISH COMMUNITIES\*

M. CARMEN PARAFITA COUTO, DIANA CARTER, PEREDUR DAVIES &  
MARGARET DEUCHAR  
*ESRC Centre for Research on Bilingualism, Bangor University*

## ABSTRACT

*Gardner-Chloros (2009) suggests that variation in code-switching (CS) can be linked to extra-linguistic factors which are either community or speaker-specific. We aim to identify which factors account for the CS patterns found in corpora we have collected of Welsh-English and Spanish-English bilingual data. We use the choice of matrix language (ML) within an MLF (Matrix Language Framework, Myers-Scotton 2002) as the dependent variable.*

*The MLF model posits that one language, the matrix language (ML), is the source of morpho-syntax in bilingual clauses. Analysing bilingual clauses in the speech of six speakers from each corpus, we find that while Welsh is uniformly the ML in clauses produced by Welsh-English bilinguals, the ML in bilingual clauses produced by Spanish-English bilinguals is varied.*

*A multivariate analysis on the Spanish-English data was conducted to test the relationship between ML distribution and extra-linguistic variables, but no significant relationship was found. However, in comparing the questionnaire data from the two corpora we argue that contrasting community-based norms may account for the difference in uniformity (Welsh-English) vs. diversity (Spanish-English) in the choice of the ML. The uniformity in the choice of the ML in the Welsh-English data can be linked to more homogeneity in self-ascribed identity and in Welsh-oriented social networks. Conversely, the variation in ML in the Spanish-English data may be related to more heterogeneity in identity and in the language of social networks.*

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

## RESUMEN

*Gardner-Chloros 2009 sugiere que factores extralingüísticos (específicos de una comunidad bilingüe o del propio hablante bilingüe) pueden ser responsables de la variación en el cambio de código (CC). El objetivo de nuestro estudio es identificar los factores que intervienen en la producción de diferentes patrones de CC en dos corpus de conversaciones bilingües (galés-inglés y español-inglés). Para el análisis aplicamos el modelo del MLF (Matrix Language Framework, Myers-Scotton 2002) y nos enfocamos exclusivamente en la variación presente en la lengua matriz (LM).*

*Según el modelo del MLF, una lengua, la LM, es la fuente morfosintáctica de la cláusula bilingüe. Al analizar las cláusulas bilingües del habla de seis hablantes de cada corpus, encontramos que, en el corpus galés-inglés, el galés constituye uniformemente la LM; mientras que en el corpus español-inglés, la LM varía entre las dos lenguas.*

*El análisis multivariante, empleado para medir la correlación entre la distribución de las lenguas matrices y las variables extra-lingüísticas, no arrojó resultados significativos. Sin embargo, al comparar los datos de los cuestionarios, argumentamos que las diferencias entre las normas de estas comunidades pudieran explicar la uniformidad (galés-inglés) vs. la diversidad (español-inglés) en la selección de la LM. La uniformidad en la selección de la LM en los datos galeses-ingleses pudiera estar relacionada a una mayor homogeneidad en la identidad y en las redes sociales orientadas hacia el galés. Por otro lado, la variación en la LM, presente en los datos españoles-ingleses, pudiera asociarse a la heterogeneidad en la identidad y el lenguaje de las redes sociales.*

Palabras clave: lengua matriz, cambio de código, español, inglés, galés

## 1. INTRODUCTION

This study is inspired by an interest in the regularities and differences that underlie language alternation in bilingual speech. In particular, we focus on naturalistic data from two bilingual corpora (Welsh-English and Spanish-English) recently collected in Wales (UK) and in Miami (Florida, USA).

We use the Matrix Language Frame (MLF) model (Myers-Scotton 2002) as a means of classifying the code-switching patterns found in our data. Analysing bilingual clauses in the speech of six speakers from each corpus, we find that while Welsh is uniformly the Matrix Language (ML) in clauses produced by Welsh-English bilinguals, there is variation in which language is the ML in bilingual clauses produced by Spanish-English bilinguals.

Using the recorded materials from our two corpora and the sociolinguistic questionnaires that were collected after each recording, we pursue the following objectives:

- (i) To analyse production data in order to assign an ML to each bilingual clause in a sample of data from each corpus.
- (ii) To perform a quantitative analysis of the results of (1) to describe the distribution of the ML in the two corpora.
- (iii) Where variation in the ML is found, to investigate the relation between the ML as a dependent variable and independent speaker-based variables.
- (iv) To use our questionnaire data to provide an overall characterisation of the two communities and to relate these to our findings regarding the distribution of the ML.
- (v) To weigh up the relative impact of speaker-based and community-based factors on the patterns found in our data.

## 2. THE MATRIX LANGUAGE FRAME (MLF) MODEL

The MLF (Myers-Scotton 1993, 2002) proposes that in ‘classic code-switching’ (Myers-Scotton 2002:8) there exists an asymmetrical

relationship between the two languages involved: one being the ‘matrix’ language and the other being the ‘embedded’ language (EL). The ML can be defined as the language which provides the morpho-syntactic frame for the clause. The EL provides inserted material (mostly content words).

### 2.1 *The two principles of the MLF model*

Myers-Scotton (2002:59) posits that there are two principles that can be used to identify the ML in a clause: the System Morpheme Principle (SMP) and the Morpheme Order Principle (MOP).

The SMP states that the ML sources outside late system morphemes, which are morphemes which 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, while the MOP states that word order will be sourced from the ML.

The MLF has been tested successfully before on both Welsh and Spanish data (cf. Deuchar, Muysken & Wang 2007; Deuchar 2006; Deuchar & Davies 2009; Davies & Deuchar forthcoming; Smith 2006).

## 3. THE BILINGUAL COMMUNITIES

The main similarities and differences between the two bilingual communities are summarised in table 1, which is partially based on Gathercole (2007).

	<b>Welsh-English</b>	<b>Spanish-English</b>
<b>Language Families</b>	Celtic, Germanic	Romance, Germanic
<b>Bilingualism since</b>	19 <sup>th</sup> century	1960s
<b>Bilingualism type</b>	Native-like	Native-like & L2 speakers
<b>Speakers</b>	Indigenous	Immigrants
<b>Speaker identity</b>	Welsh	Mixed
<b>Language use</b>	Both daily	Both daily
<b>Language visibility</b>	Both in written/spoken	Both in written/spoken

	form	form
<b>Language of cultural events</b>	Welsh	Spanish
<b>Bilingual education</b>	Since 1940s	Since 1960s

Table 1. Similarities and differences between the Spanish-English and Welsh English communities.

## 4. METHODOLOGY

### 4.1 Participants

151 Welsh-English and 85 Spanish-English participants were recorded for the two corpora. Out of those participants, recordings from a sample size of six Welsh-English and six Spanish-English speakers were chosen. The Welsh-English participants were born and raised in Wales and therefore had also been exposed to English at an early age. They reported that they had acquired both languages simultaneously or had started learning English just prior to primary school. Further details are shown in table 2.

<b>Speaker</b>	<b>Age</b>	<b>Gender</b>	<b>Education completed</b>	<b>Primary school language</b>	<b>Secondary school language</b>	<b>National identity</b>
AMR	36	Female	High school	Welsh	Welsh	Welsh
ANT	52	Female	Graduate school	Bilingual	Bilingual	Welsh
DAN	25	Male	Graduate school	Welsh	Welsh	Welsh
HEC	23	Male	Graduate school	Welsh	Welsh	Welsh
LIS	20	Female	High school	Bilingual	Welsh	Welsh
MAB	19	Female	High school	Welsh	Welsh	Welsh

Table 2. Welsh-English participants

As with the Welsh-English participants, six Spanish-English participants (from three recordings) were chosen from the main corpus (see table 3). All of the participants had acquired Spanish at two years of age, and the language spoken in the home was also Spanish. With regards to the age of acquisition of English, however, the participants showed more variability. Two of the speakers had learned English at age two, one speaker at age four, and the others had not learned English until primary or secondary school.

<b>Speaker</b>	<b>Age</b>	<b>Gender</b>	<b>Education completed</b>	<b>Primary school language</b>	<b>Secondary school language</b>	<b>National identity</b>
AME	26	Female	Bachelors	Spanish	Bilingual	Venezuelan
CAR	21	Female	High school	English	English	American
KEV	57	Male	High school	Spanish	English	Cuban
PAI	33	Female	High school	English	English	Cuban-American
SAR	34	Female	Bachelors	English	English	Cuban-American
SOF	44	Female	Graduate school	English	English	Cuban

Table 3. Spanish-English participants

#### *4.2 Data Collection*

Digital audio recordings were collected by a research team in Wales for the Welsh-English data set and in Miami for the Spanish-English set. Participants were recorded having informal conversations in pairs for approximately half an hour. In order to minimize the Observer's Paradox (Labov 1972), the investigator was not present. Recordings were afterward transcribed in full. Both corpora consist of

approximately forty hours of natural speech data. In this study a representative sample of three hours<sup>1</sup> of recordings was analysed.

In order to provide details for the extralinguistic variables examined in the current study, each speaker was requested to fill in a questionnaire, which included enquiries about gender, age, language history, and attitude towards CS.

### *4.3 Data Analysis*

All of the bilingual utterances were extracted either manually or with the Computerized Language Analysis (CLAN) programme (MacWhinney 2000). The utterances were then divided so that the resulting units of analysis were bilingual simple clauses. The data were analysed according to the MLF framework, in order to identify each clause as having either a Welsh, English or Spanish ML.

A quantitative analysis was conducted on a total of 170 bilingual simple clauses from the Welsh-English transcriptions and 149 from the Spanish-English transcriptions, using the Goldvarb X Varbrul programme (Sankoff, Tagliamonte and Smith 2005). The dependent variable was the ML of the clause, with a binary choice of either Welsh or English and Spanish or English. The independent variables were the following: gender, age, education, age of acquisition of Spanish or Welsh, age of acquisition of English, language proficiency for Spanish or Welsh, language proficiency for English, language used in primary school, language used in secondary school, language input of mother, father and guardian, CS self-report, CS attitude, national identity, and language of social network.

Our study also analysed the questionnaire responses from all of the participants for each corpus (85 Spanish-English and 151 Welsh-English) in order to reveal community-wide trends. The results from the two corpora were compared with the questionnaire responses from the sub-groups in order to ensure that the sub-groups were indeed representative of the larger groups.

## 5. RESULTS

The MLF analysis of the simple bilingual clauses from both datasets yielded the following results. For the Welsh-English data, 100% of the clauses had Welsh as the ML. The Spanish-English data, however, provided more variation: 66% of the clauses had Spanish as the ML and the remaining 34% were identified as having an English ML. Table 4 below presents the ML distribution for each Spanish-English participant.

<b>File source</b>	<b>Speaker</b>	<b>Spanish ML</b>	<b>English ML</b>
Sastre1	SOF	63%	37%
Sastre1	KEV	66.7%	33.3%
Zeledon1	AME	100%	0%
Zeledon1	CAR	93.3%	6.7%
Herring10	PAI	75%	25%
Herring10	SAR	48.5%	51.5%

Table 4. Distribution of matrix language for Spanish-English per participant

Given that 100% of the clauses from the Welsh-English data had Welsh as the ML, and therefore were completely invariant as a dependent variable, the multivariate analysis was only conducted with the Spanish-English dataset. The results did not reveal any of the extra-linguistic factors as being significant.

In order to observe corpus and community-wide sociodemographic trends, an analysis of the 85 Spanish-English and 151 Welsh-English questionnaires was also conducted. We examined the same sixteen questions that were analysed for the two sub-groups. The findings show that the age of acquisition of Spanish or Welsh is similar in that the majority of our participants started learning the language at two years of age (91% of Spanish-English and 83% of Welsh-English participants for the corpora and 100% for both sub-groups). The age of acquisition of English, however, was varied across ages for both language pairs, as seen in figure 1.

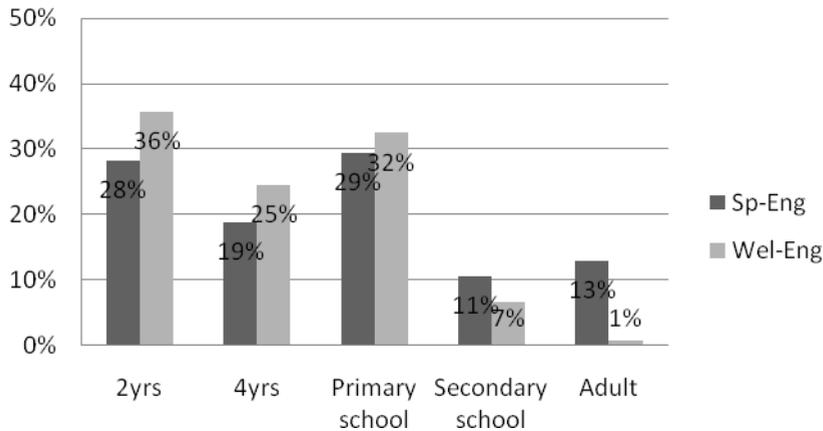


Figure 1. Age of acquisition of English for the Spanish-English and Welsh-English corpora

Both groups reported a similar proficiency in Spanish or Welsh: 74% of the Spanish-English participants and 77% of the Welsh-English participants indicated the highest level of proficiency for their questionnaire response. The results for proficiency in English revealed differences between the groups. 85% of the Spanish-English participants elected the highest proficiency in English, whereas only 68% of the Welsh-English participants reported the same level. This may be due to the differences we find in the primary and secondary school language mediums (see figures 2 and 3). The majority of the Welsh-English participants (72%) had their primary school education in Welsh while the majority of the Spanish-English group (52%) attended school conducted in English.

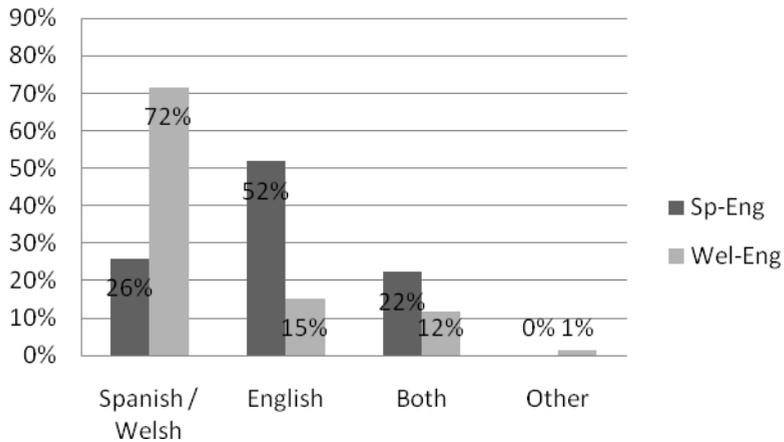


Figure 2. Primary school language for the Spanish-English and Welsh-English corpora

Further differences can also be seen by comparing the language used in secondary education. While 60% of the Spanish-English participants received their education in English, 28% of the Welsh-English participants reported the same.

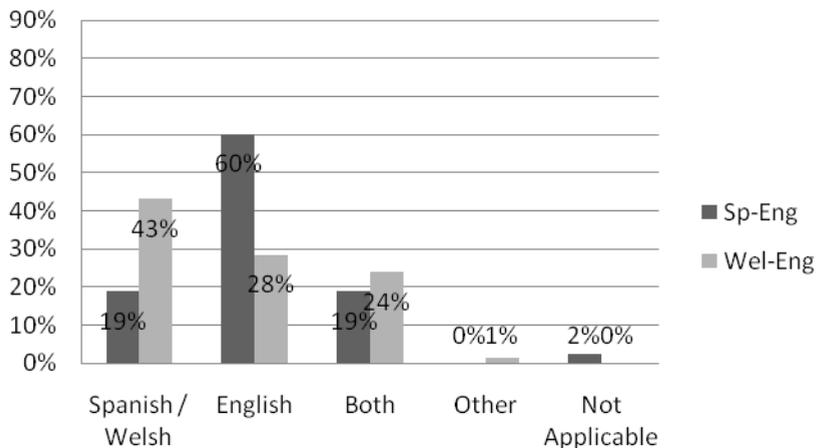


Figure 3. Secondary school language for the Spanish-English and Welsh-English corpora

Notable differences between the two communities are also clearly visible with respect to their National Identity and their Social

Networks. Almost all of the Welsh-English participants identify themselves as Welsh (90%) and half of the Spanish-English participants (48%) fall into the “other” category, which includes identities such as Venezuelan, Dominican, and Cuban-American. The remaining Spanish-English identified as American (32%) and Cuban (21%). Similar results were found with the two sub-groups, where 100% of the six Welsh-English participants elected Welsh as their identity and 50% of the Spanish-English group were classified in the “other” category. The mean scores from the Social Network analysis revealed that the Spanish-English community has a bilingual social network that includes a balance of Spanish and English. The Welsh-English participants tend to have a predominantly Welsh social network. These trends were similar for the corpora and the two sub-groups.

## 6. DISCUSSION AND CONCLUSION

In this paper we set out to identify whether speaker-based variables or community-wide norms would account for the CS patterns found in the two corpora. We found that there was considerable uniformity in the Welsh-English data (i.e. the ML was almost always Welsh) and this ruled out the possibility of using speaker-based variables to account for differences between speakers. In the Spanish-English data there was more variation in the choice of ML, allowing for the possibility that this could be related to speaker-based variables. However, none of these turned out to be significant.

We also considered the role of community-wide norms in determining (a) the uniformity of the ML in the Welsh-English community and (b) the relative flexibility in the choice of the ML in the Spanish-English community. The main differences between the two communities as shown by the questionnaire results were that (i) self-perceptions of identity were more homogeneous in the Welsh-English community, which may be attributed to Welsh-medium education (cf. figures 2 and 3) and (ii) social network was oriented more to one language (Welsh) in the Welsh-English community whereas close contacts in the Spanish-English community were more balanced between the two languages. The fact that the Welsh-English

speakers consider themselves to have a mainly Welsh identity and also address their closest contacts mostly in Welsh could account for their virtually exclusive choice of Welsh as a matrix language. On the other hand, the more variable identity and choice of language with close contacts in the Spanish-Welsh community could account for the more variable choice of both English and Spanish ML.

It is also possible that language-internal factors may play a role. There may be a universal tendency to select one ML as suggested by Chan (2009). It could be argued that it would be more parsimonious for bilinguals to choose one language as the ML unless the two languages concerned have similar word order. While Spanish and English have similar order (SVO), the word order of Welsh (VSO) and English (SVO) contrasts. Hence, inter-clausal switching between a VSO and an SVO language may be dispreferred. If this is the case, Welsh-English speakers will tend to choose just one ML, and as suggested above, it may be their Welsh national identity and Welsh-oriented social network that determines Welsh as the one they choose. For Spanish-English speakers, however, the fact that both languages have SVO word order allows for more flexibility in the choice of ML. Nevertheless, future research will be needed in order to isolate the role of structural factors of this kind, for example by studying a community where structural and social factors point in the opposite direction. This would occur if a community were to be like the Welsh-English one in using two structurally different languages but unlike the Welsh-English in showing less homogeneous identity and less orientation to Welsh social networks.

## NOTES

\* We are grateful to Fraibet Aveledo, Marika Fusser, Jon Herring, Siân Lloyd-Williams, Elen Robert, Alberto Rosignoli and Jonathan Stammers for their comments and help collecting and transcribing the data.

<sup>1</sup>This is due to the fact that only three conversations (i.e.90 minutes total) were transcribed in the Spanish-English corpus at the time this research was conducted. For the sake of comparison, a similar amount of data was chosen from the Welsh-English corpus.

## REFERENCES

- Chan B. 2009. "Code-switching between typologically distinct languages". In ed. Bullock, B. and Toribio, A.J. *The Cambridge Handbook of Linguistic Code-switching*. Cambridge University Press.
- Davies P., Deuchar M. Forthcoming. "Using the Matrix Language Frame model to identify word order convergence in Welsh-English bilingual speech", in ed. Breitbarth, A., C. Lucas, S. Watts and D. Willis., *Continuity and Change in Grammar*. Amsterdam: John Benjamins
- Deuchar M. 2006. "Welsh-English codeswitching and the Matrix Language Frame model", *Lingua*, 116 (11): 1986-2011.
- Deuchar M., Davies P. 2009. "Code-switching and the future of Welsh", *International Journal of the Sociology of Language*, 195: 15-38.
- Deuchar M., Muysken P., Wang S. 2007. "Structured Variation in Codeswitching: Towards an Empirically Based Typology of Bilingual Speech Patterns", *The International Journal of Bilingual Education and Bilingualism*, 10(3): 298-340.
- Gardner-Chloros P. 2009. *Code-switching*. Cambridge, UK.
- Gathercole V. 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(3):224-247.
- Labov W.1972. "Some principles of linguistic methodology". *Language in Society*, 1 (1), 97-120
- MacWhinney B. 2000. *The CHILDES Project: Tools for Analyzing Talk (3<sup>rd</sup> ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Myers-Scotton C. 1993. *Duelling languages: Grammatical structure in codeswitching*. Oxford and New York: Oxford University Press.
- Myers-Scotton C. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford and New York: Oxford University Press.
- Sankoff D., Tagliamonte S. A., Smith E. 2005. *Goldvarb X*. Toronto, Canada: Department of Linguistics, University of Toronto.

Accessed on 25 January 2009 at  
[http://individual.utoronto.ca/tagliamonte/Goldvarb/GV\\_index.  
htm](http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm)

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