



Canolfan ESRC Centre  
dros Ymchwil for Research  
i Ddwyieithrwydd on Bilingualism

# Welsh-English and Welsh-Spanish code-switching: a preliminary comparison

Diana Carter, Peredur Davies, Margaret Deuchar, Kevin  
Donnelly, Lowri Jones, M. Carmen Parafita, Myfyr Prys



Cyngor Cyllido Addysg  
Uwch Cymru  
Higher Education Funding  
Council for Wales

hefcw



# Acknowledgements ...



- ▶ AHRC
- ▶ ESRC
- ▶ Marika Fusser
- ▶ Siân Lloyd
- ▶ Elen Robert
- ▶ Gary Smith
- ▶ Jonathan Stammers

## ▶ **Welsh-English**

- ▶ 40 hours of conversation by 151 speakers
- ▶ Collected 2005-2007 in North Wales
- ▶ **Talkbank**
- ▶ draft web interface: **BangorTalk**
- ▶ Davies6

## ▶ **Welsh-Spanish**

- ▶ 20 hours of conversation by 92 speakers
- ▶ Collected October-November 2009 in Patagonia
- ▶ initial output: **BangorTalk**
- ▶ Patagonia2

# Data transcription

- ▶ CHAT system (TalkBank, MacWhinney 2000)
- ▶ Main tier
- ▶ Translation tier
- ▶ Word-by-word glosses
- ▶ Our new approach to glossing: autoglossing

# Autoglossing aims

- ▶ Save time on glossing the text
- ▶ Improve access to the corpora
- ▶ Longer-term: more versatile statistical analysis

- ▶ Process:
  - ▶ Import transcribed texts into a database
  - ▶ Look up each word in a dictionary to get POS tags
  - ▶ Disambiguate homonyms using CG (110 rules for Welsh)
  - ▶ Print out the autoglossed texts
- ▶ Handles both manually-glossed and unglossed texts
- ▶ One-click unattended import (6-7 minutes/text)
- ▶ Re-uses existing free (GPL) resources

# BangorTalk

*Conversational corpora from the ESRC Centre*

## *Welcome to BangorTalk*

### The ESRC Centre

This is a test site for the conversational corpora assembled by the **ESRC Centre for Research on Bilingualism in Theory & Practice** at University of Wales Bangor.

We are seeking to gain a greater understanding of how bilingual individuals in a variety of communities manage both their languages within the same conversation. The questions we consider include:

1. Do bilinguals in different types of communities handle their two languages in different ways in conversation?
2. How do social variables such as class, age and gender affect the way people handle their two languages in informal conversations?

### The corpora

To date, we have assembled three corpora:

### Change language

(not functional yet)

English (Saesneg) ▾

### Contact us

bilingualism@bangor.ac.uk

### The corpora

**The Siarad corpus**  
**The Patagonia corpus**

# Browser interface

- ▶ Databased texts offer more presentation options
- ▶ Accessible from anywhere on the web
- ▶ No need to install CLAN software
- ▶ Queries similar to CLAN, but via browser
- ▶ **DANGER! WORK IN PROGRESS!**

# BangorTalk: text presentation

## Glossed version of *Siarad corpus: deuchar1*



Click the play button to hear this conversation.

[Download the audio.](#)

[View the chat file.](#) (To download, right-click and choose *Save Link As*.)

The first line below shows each utterance in **deuchar1** as transcribed, containing pause and expression markers. The second line gives the text with those markers and all elisions removed - moving your mouse over the individual words in this line will give a popup with part-of-speech information. The last line gives an English rendition of the text.

- 1 MYF um@0 # oedd@1 [/?] oedd@1 o@1 wedi@1 setio@1 pethau@1 fyny@1 i@1 ni@1 # er@0 copio@1 .  
um oedd oedd o wedi setio pethau fyny i ni er copio .  
*um, he'd set things up for us to copy NONFIN*
- 2 MYF um@0 # oh@0 still@0 life@2 mae@1 nhw@1 (y)n@1 galw@1 o@1 [=? fo@1] .  
um oh still life mae nhw yn galw o .  
*um, oh they call it still life*
- 3 SER +< oh@0 ia@1 .  
oh ia .  
*oh yes*
- 4 MYF um@0 glass@0 hir@1 fel@1 (y)na@1 # efo@1 fel@1 glass@0 baubles@2 yno@1 fo@1 yn@1 y@1 top@0 .  
um glass hir fel yna efo fel glass baubles yno fo yn y top .  
*um, a long glass like that with, like, glass baubles in it in the top*

### Change language

(not functional yet)

English (Saesneg) ▾

[The Siarad corpus](#)  
[The Patagonia corpus](#)

[deuchar1 menu](#)

# BangorTalk: word frequency

## Word frequency for Patagonia corpus: patag

### Words with language tag cy (4379)

**yn** (stative STAT) [245], **i** (to P) [192], **a** (and C) [147], **oedd** (be VFLE 3S IMPERFECT) [133], **mae** (be VFLE 3S PRESENT) [122], **yr** (the T) [107], **ia** (ia UNK) [90], **y** (the T) [85], **hi** (she) [83], **o** (from P) [82], **yn** (stative STAT) [74], **dw** (dw UNK) [72], **ie** (yes B) [64], **yn** (in P) [62], **ddim** (zero N SM S) [60], **i** (I R) [57], **nhu** (but C) [47], **oeddwn** (be VFLE **yna** (there B) [45], **wedyn** (afte (not\_(negative\_responsive\_verl **mynd** (go VINF) [39], **efo** (with [36], **wel** (wel UNK) [33], **ni** (we (know\_(fact) VINF) [27], **ar** (on (very B) [23], **ysgol** (school N F

## Instances of "byth" in Patagonia corpus: pa

- 36 AVR pob tro dw i (y)n mynd i Esquel@s:cy&es dw i (y)n meddwl am Lea@s:cy&es ond byth yn ei gweld hi .  
pob tro dw i yn mynd i Esquel dw i yn meddwl am Lea ond byth yn ei gweld hi .  
*every time I go to Esquel I think of Lea, but never see her .*
- 70 CHT o(eddw)n i (.) byth yn cael atebiad wrthyn nhw .  
oeddwn i byth yn cael atebiad wrthyn nhw .  
*I never got a reply from them .*
- 102 CHT a byth ar\_ôl hynny dw i ddim yn llawer am y môr .  
a byth ar\_ôl hynny dw i ddim yn llawer am y môr .  
*and ever since then I've not been keen on the sea .*
- 160 AVR wna i byth anghofio .  
wna i byth anghofio .  
*I shall never forget .*
- 162 AVR wna i byth anghofio (y)r [//] yr [/] yr [/] yr er@s:cy&es +. .  
wna i byth anghofio yr yr er .

# Bilingual corpora



Canolfan ESRC Centre

Research  
Bilingualism

## Spanish-English

**Siempre está *promising* cosas**

( Mae bob amser yn addo pethau  
He's always promising things )



Sbaeneg  
Spanish

Saesneg  
English

## Welsh-English

**Dw I'n *love-i*o soaps**

( Dw I'n caru operaau sebon  
I love soaps )



Cymraeg  
Welsh

Saesneg  
English

**Tocarwch y *vicina***

( Garwch y corn  
Hoot the horn )



Sbaeneg  
Spanish

Cymraeg  
Welsh

## Spanish-Welsh

# Bilingual conversation

## Wales and Patagonia



### Wales

---

Mainly Welsh used

---

English words in Welsh sentences, but Welsh grammar is used

---

Young and old speakers highly fluent

---

Parents speak Welsh to children

### Patagonia

---

Both Welsh and Spanish used

---

Few Spanish words in Welsh sentences

---

Younger speakers less fluent in Welsh than older speakers

---

Parents rarely speak Welsh to children

- ▶ 151 Welsh-English and 92 Spanish-Welsh speakers
- ▶ Conversations recorded in pairs
- ▶ Researcher not present for the conversation
- ▶ Background recorded via questionnaire
- ▶ Conversation transcribed using CHAT
- ▶ Sample of recordings from 12 speakers (6 Welsh-English, 6 Spanish-Welsh) selected for analysis

# Preliminary analysis: objectives

- ▶ Compare CS patterns in the Welsh-Spanish and Welsh-English communities using the Matrix Language Frame model
- ▶ Examine the relationship between CS patterns and extralinguistic community characteristics

- ▶ We extracted all simple clauses from the transcriptions
  - ▶ 1,862 Welsh-English clauses
  - ▶ 1,099 Spanish-Welsh clauses
- ▶ We assigned a ML to each clause

# Matrix Language Frame model

- ▶ In “classic CS” (Myers-Scotton 2002) there is an asymmetric relationship between the two languages in a clause:
  - ▶ One is the matrix language (ML)
  - ▶ The other is the embedded language (EL)
- ▶ ML provides the morphosyntactic frame for the clause
- ▶ EL provides mainly inserted material (mostly content words)

- ▶ **Morpheme Order Principle (MOP)**
  - ▶ Word order is sourced from the ML
- ▶ **System Morpheme Principle (SMP)**
  - ▶ The grammatical morphemes are sourced from the ML

**mae**            **Americans yn mwy commercial**  
be.PRES.3S    Americans    PRT    more    commercial

*“Americans are more commercial.”* (Fusser 27:430)

MOP: **mae** precedes the subject **Americans** (VS)

SMP: **mae** has Welsh finite subject-verb (anti)agreement

**bydda**      **i (y)n wneud biotechnoleg, [achos**  
be.FUT.1SG | PRT do.INF biotechnology because  
**dw**            **i ddim yn gallu inscribir=me**            **yn**  
be.PRES.1SG | NEG PRT can.INF enrol.INF=ACC.1SG in  
**ddau]**  
two

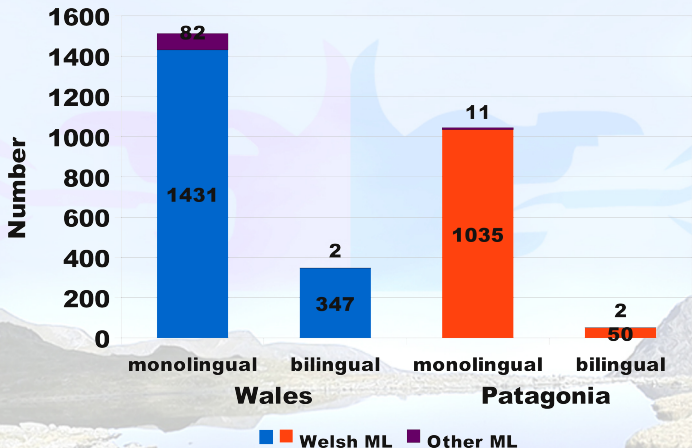
*"I'll be doing biotechnology, because I can't enrol in both."*  
(Patagonia 29)

MOP: **bydda** precedes the subject **i**

SMP: **bydda** has Welsh finite subject-verb agreement

# Clause analysis

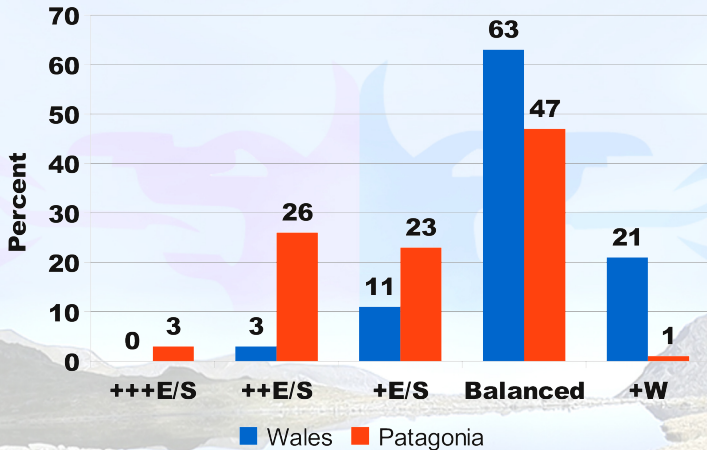
Type of clause, and Matrix Language



# Proficiency in minority language



# Bilingual proficiency



- ▶ Comparison of CS patterns in the Welsh-English and Welsh-Spanish communities:
  - ▶ Welsh is main ML in both
  - ▶ More bilingual clauses (intrasentential CS) in the Welsh-English data
- ▶ Key extralinguistic variables showing a link between CS patterns and community norms:
  - ▶ Proficiency in minority language
  - ▶ Degree of balanced bilingualism

# Selected references 1

**Chan, B. (2009).** *Code-switching between typologically distinct languages.*

In: D. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-switching.* Cambridge: 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: A. Beitbarth, C. Lucas, S. Watts & D. Willis (Eds.), *Continuity and Change in Grammar.* Amsterdam: John Benjamins.

**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, 224-247.

## Selected references 2

**MacWhinney, B. (2000).** *The CHILDES Project: Tools for Analyzing Talk (3rd ed.)*.

Mahwah, NJ: Lawrence Erlbaum Associates.

**Myers-Scotton, C. (2002).** *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford and New York: Oxford University Press.

**Sankoff, S., Tagliamonte, S. & Smith, E. (2005).** *Goldvarb X*. Toronto, Canada: Department of Linguistics, University of Toronto.